

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:14:24 ; Search time 1037.86 Seconds
(without alignments)
6784.436 Million cell updates/sec

Title: US-09-599-087-4
Perfect score: 806
Sequence: 1 ggaacgagggaaatctgcc.....actcaatgcagacacaaaaa 806

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	593.4	73.6	906	141	BE899580 601682443
2	488.8	60.6	503	7	AA422178 zv31907.f
3	475	58.9	517	120	AM854263 RC3-C7025
4	460.8	57.2	467	102	AI831407 wj64a12.x
5	460.6	57.1	476	111	AM134688 UI-H-B11-
6	460	57.1	460	104	AI983767 wu20a04.x
7	455.2	56.5	460	142	BF001316 7g62b10.x
8	451.8	56.1	455	102	AI833391 at69f03.x
9	451.8	56.1	455	102	AI833391 at69f03.x
10	450	55.8	453	116	AM516596 xq01h07.x
11	447.4	55.5	450	104	AI948903 wq17c02.x
12	442.4	54.9	444	102	AI833297 at67a07.x
13	436.2	54.1	441	9	AA587764 nm95f05.s
14	428.4	53.2	430	18	AI283185 qk49g09.x
15	426.8	53.0	431	102	AI832498 at69e07.x
16	426.8	52.9	429	102	AI813445 wj06e01.x
17	423.6	52.6	487	9	AA535959 n101d03.s
18	419.8	52.1	423	19	AI336470 qo61d01.x

Accession	AA422178	GI:2101029
Version	AA422178.1	
Keywords	EST.	
Source	human.	
Organism	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
Reference	1 (bases 1 to 503)	
Authors	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kuwabara, T., Lacy, M., Le, N., Lennon, G., Merit, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.	
Title	WashU-Merck EST Project 1997	
Journal	Unpublished (1997)	
Comment	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28m3 rev2 ET from Amersham High quality sequence stop: 503.	
Features	Location/Qualifiers	
Source	1..503	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:755292"	
	/clone_lib="Soares ovary tumor NbHOT"	
	/sex="female"	
	/tissue_type="ovarian tumor"	
	/lab_host="DH10B (ampicillin resistant)"	
	/note="Organ: Ovary; Vector: pT7SD (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGGAGGCGGCGGCTTTTCTTTTCTTTT 3'], RI adapter-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7S3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	108 a 167 c 119 g 109 t	
ORIGIN		
Query Match	60.6%; Score 488.8; DB 7; Length 503;	
Best Local Similarity	99.4%; Pred. No.2.6e-126;	
Matches 501; Conservative	0; Mismatches 2; Indels 1; Gaps 1;	
QY	14 atctgctcttcacatgaggtcttagctcttcacagctgctgtgatacctgcttc 73	
Db	1 ATCTGCTTCTCACATGAGGCTTCTAGTCCTTTCAGGCTGCTGTATCCTGCTTTC 60	
Y	74 tggcttcacatcttcctccacagaagggaaggaggtctctgccaaggctgtgtcaggagcag 133	
Db	61 TGGCTTCTCATTTTCTCCACAGAGGGGAGAGAGGCGTCTGCCAAGCCGTGTACAGCAGG 120	
QY	134 agaacagagctctgtgccaacgagctccatgcccccaactcaacaacttgaagaagacat 193	
Db	121 AGAACCAAGGCTTGCTGCGCACCGGATCTCCCTAGCCCACTCAACAACATGTAAGAGCAT 180	
QY	194 catgtgaggtctgttaaacatgcaagcttgaagccagagccccgcgcttgyggtgagcct 253	
Db	181 CATGTGAGGCTCTGTAAACATGCAAGCTTGTAGCCAGAGCCCCGCTTGTGGTGTGCT 240	
QY	254 ggggacccccaaggtgttagacatcccaaggcaagactccagaaaggggaaacctaat 313	
Db	241 GGGG-AGTCCCAACAGGTGTAGCATTCCCAAAACAAGACTCCAGAGAGCGGAAGACCTAT 299	
QY	314 gcttgacacctgaggtaaccagcaagctctctctcccttcaagcttcaagcagatga- 373	
Db	300 GCGTGGCACTGAGGTACCCAGACAGACCTCTCTCTCCCTTTACAGCTTACACAGCATGA 359	

Qy	374	gctgcgaatttggaaggagctcattccgggctcgaagagccctgggaaagttccaaactc	433
Db	360	gctgcgaatttggaaggagctcattccgggctcgaagagccctgggaaagttccaaactc	419
Qy	434	caagctctctgtctcaattgtgcatcaacttcacagagctatcatgagccaactcacc	493
Db	420	cacgacgcttctgctcaattgtgcatcaacttcacagagctatcatgagccaactcacc	479
Qy	494	acagagcgctcagtcgccaacatgt	517
Db	480	acagagcgctcagtcgccaacatgt	503
RESULT	3		
LOCUS	AM854263	517 bp	EST
DEFINITION	RC3-CT0254-100500-211-g12	CT0254	Homo sapiens cDNA, mRNA sequence.
ACCESSION	AM854263		
VERSION	AM854263.1	GI:7949956	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	EnuAryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Bionesi, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordino, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7)	3491-3496 (2000)
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?file=6t2-RC3-CT0254-100500-211-g12&t3=2000-05-10&t4=1) Seq primer: puc 18 forward High quality sequence stop: 6 High quality sequence stop: 517. Location/Qualifiers		
FEATURES			
SOURCE	1..517		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_id="CT0254"		
	/dev_stage="Adult"		
	/note="Organ: colon; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	106 a	172 g	121 t
ORIGIN	118 c	172 g	121 t
Query Match	58.9%	Score 475;	DB 120; Length 517;
Best Local Similarity	98.2%	Pred. No. 1,9e-122;	
Matches 502; Conservative	0;	Mismatches 5;	Indels 4; Gaps 2;
246	tggtgctctgggggcaactcccaagaagtgtgtgacatctcccaagaagcaagacttcagacgcgag	305	


```

Db 517 TGGTGGCTGGGGACACGCCAGGTGAGACACTCCAAACAGACTCCAGACAGCGAG 458
Qy 306 aacctatgctgagcactgaggtaccagagcctctctctcccttcaagcctcac 355
Db 457 AACCTATGCTGGCA--TGAGGTTCCAGACAGCTCTCTCTCTCTCTCTCTCTCT 400
Qy 366 agcagtgagctgagctgaggtgaggttactctcggctgagcagcctgggaagttc 425
Db 399 AGCAGTGAAGTGAATGTGGAGGCTTCACTCTGGGCTGCAAGGACCCCTGGAAAGTTC 340
Qy 426 cagaatccagctcctctctcattgtgcatcaattcagagctatcatgagccaac 485
Db 339 CAGAACTCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 280
Qy 486 ctaccccaagaggctcagtcgacacatgtgagctctcagtgcaaacacagcaga 545
Db 279 CTCACCCACAGAGGCTCTAGTCCGACACAGTGGGCTCTCCAGTGCAACCGAGCA 220
Qy 546 ttccacatgagccggtcacaagctacaataccagagacatcatcctgctagagtcag 605
Db 219 TTCACCAAGACCGGTGACAGCTACAAATCCAGACACATCAATCTCTGAGAGTGACAG 160
Qy 606 gtggcaagaccacaaaggtggtctgacacagactgacagagctcctcatctttagtca 665
Db 159 GTGGCAAGACCCAGAGGTGCTGACCAAGACTGCAAGAGTCTCTCATCTTAGGTCCA 100
Qy 666 ttaagcctctgcatcttaactacagcagtcagtcgcccaggaatcctctctag 725
Db 99 TTCAGCTCTCTGGCATTACTACGACATCCAGTG--TCCCAAGAGATCCCTTCTAGC 42
Qy 726 ctctgacatgagctgctgctggaagagcgc 756
Db 41 CTCCTGACATGAGTCTGTGAAAGAGCATC 11

RESULT 4
AT831407/c 467 bp mRNA EST 21-DEC-1999
LOCUS WJ6412.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407582 3'
DEFINITION mRNA sequence.
ACCESSION AI831407
VERSION AI831407.1 GI:5452078
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 467)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnli.gov/dbrrp/image/image.html
Insert length: 756 Std. Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
Location/Qualifiers
1..467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2407582"
/clone_lib="NCI_CGAP_Lu19"

```

```

/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pTZ193D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pTZ193 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 95 a 94 c 146 g 132 t
ORIGIN

Query Match 57.2%; Score 460.8; DB 102; Length 467;
Best Local Similarity 99.6%; Pred. No. 1,8e-118;
Matches 462; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 343 ctgtctcccttcacgcttcacagcagtgagctgcaatgttgagggcttaccggy 402
Db 467 ctgtctcccttcacgcttcacagcagtgagctgcaatgttgagggcttaccggy 408
Qy 403 ctgcaaggaacctgggaaggttcacgaactcacagctctgtctcaattgtgcatcaac 462
Db 407 CTCGAAGAGACCCGGGAAAGTTCCAGAACTCCAGCTCTCTCTCTCTCTCTCTCTCT 348
Qy 463 ttcaagagctatcatgagcacaacctaccccaagagctcagtcgacacatgttgggc 522
Db 347 TTCAGAGCTATCATGAGCAACCTCACCCAGAGGCTCTCATGTCACACAGTGGGCC 288
Qy 523 ttcccaagtgcacaaccgagacatccacatgacccggtcacaagctacaatcagaagac 582
Db 287 TCTCCAGTGCACAAACCCCGACAGCTTCCACCATCAGCGGTGACAGTCAAAATCCAGAGAC 228
Qy 583 catcaatcctgtctagagtgcaaggtgcaagcaccacaaaggtgtgctgacacaaagctgcag 642
Db 227 CATCATCTCTGCTGAGAGTGACAGGAGGCAAGCAACCCAAAGGTGCTGACCAAGACTGCGAC 168
Qy 643 agtctcctcatcttcagtcacatcagcctcctgcatcttaactacacagcatcagttg 702
Db 167 AGTCTCTCTCATCTTCCAGGTCATGACCTCTCTGACCTCTGACATTTAACTACAGATCCAGTGG 108
Qy 703 ttcccaaggaatcccttctagctctcagctctgacatgattctgctggaagagcatccaaac 762
Db 107 TCCCAAGGAATCCCTTCTAGCTCTCTGACATGATGTGCTGGAAGAGCATCCAAACA 48
Qy 763 aacaagtaataataataataataataataataataataataataataataataata 806
Db 47 AACAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4

RESULT 5
AW134688 476 bp mRNA EST 28-OCT-1999
LOCUS UI-H-B11-bdp-f-09-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
DEFINITION IMAGE:271213 3', mRNA sequence.
ACCESSION AW134688
VERSION AW134688.1 GI:6138234
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 476)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

```


Db 167 GGTGACCAAGACTGACAGTCTCTCCACCTTCAGTTCATTCAGCCTCTGAGCATTTA 108
Qy 685 actaacgacatccagtgctgcccccaagaatcccttcttaagcttccctgaatgctgct 744
Db 107 ACTACAGACATCCAGTGTGCTCCCAAGGAATCCCTCTTACCTCTTACATGAGTGTCT 48
Qy 745 ggaagaagcat-ccaacaacaagaataataataataaactc 750
Db 47 GGAAGAGCATACAAACAACAAGTATATATATATATATATATATATATATATATAT 1

RESULT 10
AM516596/c 453 bp mRNA EST 03-MAR-2000
LOCUS IMAGE:2748637 3', mRNA sequence.
DEFINITION XQ01007.X1 Soares,NHCEC_cervical_tumor Homo sapiens cDNA clone
ACCESSION AM516596
VERSION AM516596.1 GI:7154678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 453)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Glibco
High quality sequence stop: 451.

FEATURES
source
1..453
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2748637"
/clone_lib="Soares,NHCEC_cervical_tumor"
/issue_type="tumor"
/lab_host="DH10B (phage-resistant)"
/note="Organ: cervix; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dt) primer [5'
GCTTACCAATCTGAAGTGGAGCGCGCCGAGAGTATTTTATTTTATTTTATTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
Library is normalized; constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 95 a 92 c 142 g 124 t
ORIGIN

Query Match 55.8%; Score 450; DB 116; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.9e-115;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 343 ctgtctcccttccagcttccacagcagtgagctgcaatgttgagggtcctcatcggg 402
Db 453 CTGTCTCCCTTTCAGCTTCACAGAGTGAAGTGCATTTGAGGGCTTCATCTCGGG 394
Qy 403 ctgcaagagaccctgggaaatctcagaactcagctcctgtctcaattgtgccatcaac 462
Db 393 CTGCAAGAGACCTTGGGAAATCTCCAGAACCTCCGCTCTGCTCAATTTGCCATCAAC 334
Qy 463 ttccaagatcatcatagcacaactcaccaccaagggtcctcaatgcaccaatgtgggc 522
Db 333 TTTCAGAGCTATCATGATGACCAACCTCACAGGAGGCTCACTGCGCACATGTGGGCC 274
523 ttcccaagtcaaacacacagcagcttccacatgacccggtcaacaataccagaagac 562

Db 273 TCTCCAGTGCAGAACACACGAGCATTCACCATGACGGGTACAGCTCAATATCCAGAGAC 214
Qy 583 catcaatccctgtagagtgcaagggtgcaagcaccaccaagggtgctgcaagaactgcag 642
Db 213 CATCAATCTCTGTAGATGAGTGCAGGGTGGCAAGCCCAAGGGTGGCTGACCAAGACTGAG 154
Qy 643 agtctcctcatcttcagtgccattcagcctcctcgtgattactaccagcatccagtg 702
Db 153 AGTCTCTCATCTTCAGTGTCCATTCAGCCTCTGAGCATTTAACTACACATCCAGTGG 94

Qy 703 tccccaagatcccttcagcctcctcagcctcctgacatgctgctggaagacatccaaca 762
Db 93 TCCCAAGAGAAATCCCTTCTTACGCTCTCTGACATGAGTGTGCTGGAAGACATCCAAACA 34
Qy 763 aacaagtaataataataaactcaaa 792
Db 33 AACAGTAAT 4

RESULT 11
A1948903 450 bp mRNA EST 08-MAR-2000
LOCUS WQ17C02.X1 NCI-CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2471522 3',
DEFINITION mRNA sequence.
ACCESSION A1948903
VERSION A1948903.1 GI:5741213
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 450)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmer-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbip/image/image.html
Insert Length: 607 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 438.

FEATURES
source
1..450
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2471522"
/clone_lib="NCI-CGAP_Kid12"
/issue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site: 2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 94 a 92 c 141 g 123 t
ORIGIN

CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www.bio.linn.gov/bdrp/image/image.html
 Insert Length: 904 Std Error: 0.00
 Seq primer: -40m13 fwd. EM from Amersham
 High quality sequence stop: 426.

FEATURES

source

1. .441

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1076001"

/clone_1lb="NCI-CGAP_C09"

/issue_type="colon tumor RER+"

/lab_host="DH10B"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER+ colon tumor, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares4)

BASE COUNT

89 a 94 c 137 g 121 t

ORIGIN

Query Match

54.1%; Score 436.2; DB 9; Length 441;

Best Local Similarity 99.3%; Pred. No. 1.4e-111;

Matches 438; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 357 agcttcacagcaatgagctgcaatgttggaggcttcacatcggtcgtaaggaacctg 416
 Db 441 AGCTTCACAGCAATGAGCTGCAATGTGGAGGCTTCATCTCGGCTGCAAGACCTG 382
 QY 417 ggaagttccagaactcagctcctgttctcaattgtgcatcaacttcagagatca 476
 Db 381 GGAAGTTCCAGAACTCAAGCTCCTGTGCTCAATTGTGCACTCAACTTCAGACTATCA 322
 QY 477 tgaagcaactcaccacagagcctcagtcgcaacatgtggccttcacagtgcgaac 536
 Db 321 TGAGCCAACTCCACCCACAGGCGCTCAGTCGCACCATGTGGCGCTCTCCAGTGAAC 262
 QY 537 caccgagcatccaccatgaccgttcacagttacaatccagaagacatcatctgcta 596
 Db 261 CACGAGCATTCACATGACCGGTCAAGCTCAAAATCCAGAACCATCAATCTGTCTA 202
 QY 597 gagtgcaggtggtgcaagcaacccaggggtgtgtgcaagagctgagagctccatcat 656
 Db 201 GAGTGCAGGAGGCAAGCAACCCAGGGGTGCTGCAAGAGTGCAGAGTCTCTCCGTCT 142
 QY 657 tcaagttccatcagcctcctggtcatttaactaccagcatcagctggtccccaagaaacc 716
 Db 141 TCAGGTCCATTCAGCTCTGCTGCAATTAACATCCAGATCCAGTGTGCCCAAGAAATCC 82
 QY 717 ctctctgctcctcgaatgagctgtgtggaagagcatccaaacaaagttataaat 776
 Db 81 CTCTCTGCTCCTGCAATGAGTGTGCTGGAAGAGCATCCAAACAAAGTATATAAT 22
 QY 777 aaataaataactcaatgcag 797
 Db 21 AAATTAATTAATCAATGCAG 1

RESULT 14

AI283185/c

LOCUS AI283185 430 bp mRNA EST 28-JAN-1999
 DEFINITION gk49g09.x1 NCI-CGAP_C08 Homo sapiens cDNA clone IMAGE:1872352 3',
 mRNA sequence.

ACCESSION AI283185

VERSION

AI283185.1 GI:3921418

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

1 (bases 1 to 430)

TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www.bio.linn.gov/bdrp/image/image.html
 Insert Length: 933 Std Error: 0.00
 Seq primer: -40NP from Gibco
 High quality sequence stop: 403.

FEATURES

source

1. .430

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1872352"

/clone_1lb="NCI-CGAP_C08"

/issue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

88 a 90 c 133 g 119 t

ORIGIN

Query Match

53.2%; Score 428.4; DB 18; Length 430;

Best Local Similarity 99.8%; Pred. No. 2.2e-109;

Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 361 ttacagcagtgagctgcaatgttggaggcttcacatcggtcgtaaggaacctggaa 420
 Db 430 TTACAGCAGTGAGCTGCAATGTGGAGGCTTCATCTCGGCTGCAAGACCTGGGA 371
 QY 421 agttcagaactcagctcctgttcaattgtgcatcaacttcagagctatcag 480
 Db 370 AGTTCAGAACTCCAGCTCTGCTGCAATGTGCTCACTCACTTCAGAGCTATCATGAG 311
 QY 481 ccaacccaccaccccaagggctcagtgccacatctgtgggctctccagtgcaaacacc 540
 Db 310 CCAACCTCACCCCAAGGGCTCAGTGCACCATGTGGGCTCTCCAGTCAACACACC 251
 QY 541 gagcattccacacagcagctgacagcttacaacacacagagacacacacacacacac 600
 Db 250 GAGCATTCACACAGCAGCGGTGCAAGTCAATCCAGAGACATCAATCTGCTAGAGT 191
 QY 601 gcaggttgcaagcaccacaaaggtgtgtgacacagactgcagagctctccatctcag 660
 Db 190 GCAAGGAGGCAAGCAACCAAGGCTGTGACCAAGACTGCAGAGTCTCTCCATCTCAG 131
 QY 661 gtccattcagctcctcgtgcatcttaactacacagacatcagtggtcccaaggaatccctc 720
 Db 130 GTCCATTCAAGCTCTCTGCAATTAACATCCAGATGAGTCCCAAGGAATCCCTTC 71

OY	721	ctgcctccctaccacatgagctgtcgtcgagaagagcatcccaacaacaagtaataaat	780
Dd	70	CTAGCCTCTGACATGTGGTCTGCTGGAAAGGCATCCAAACAACAAGTAATAAATAAT	11
OY	781	aataaacctc 790	
Dd	10	AAATAAACTC 1	
RESULT	15		
LOCUS	A1832498/c		
DEFINITION	A1832498	431 bp	mRNA
VERSION	AI696.007.x1	Barsstead colon HPIRB7	Homo sapiens CDNA clone
KEYWORDS	A1832498.	AI832498.	GI:5454478
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Human sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	1 (bases 1 to 431)		
COMMENT	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,		
	Kitzman,D., Kucabada,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,		
	'J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,		
	White,Y., Wylie,T., Waterston,R. and Wilson,R.		
	WashU-NCI human EST Project		
	Unpublished (1997)		
	Contact: Willson RK		

```

FEATURES
source
Email: est@watson.wustl.edu
This clone is available royalty-free through IHLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from GIDCO.
Location/Qualifiers
1..431
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2377284"
/clone_lib="Barstead colon HPLRB7"
/sex="male"
/_sex="male"
/_dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note="Organ: colon. Vector: p7R3D-pac (Pharmacia) with a
modified polylinker. Site.1: EcoRI; Site.2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACGAATCTGAAAGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5' ATTACTACTACTATG 3' and 5' ATTACTACTG 3']. digested
with Not I and cloned into the Not I and Eco RI sites of
the modified p7R3 vector. Library constructed by Bob
Barstead."
89 a 92 c 142 g 107 t 1 others

```

Query Match	53.0%	Score 426.8	DB 102	Length 431
Best Local Similarity	99.3%	Pred. No. 6.1e+109		
Matches 428	Conservative	0	Mismatches	3
			Indels	0
			Gaps	0

QY	342	ccctctccccccttcaagccttcaagcagcagcagtcgaatgttgaaggttcacacccg	401
Db	431	ccgtctccccccttcaagccttcaagcagcagcagtcgaatgttgaaggttcacacccg	373
QY	402	gctctcaagcagccttggaaagttccagaactcagcgtccttgtctcaattgtgcaccaa	461
Db	371	gctctcaagcagccttggaaagttccagaactcagcgtccttgtctcaattgtgcaccaa	312
QY	462	cttccagagctatcatgagccaactcaccccaagagcctcagtcgcacacatgttggc	521

Db	311	CTTTCAGAGCTATCATATGAGCCAACTCCACGCTCAGCCGACAGCATGTGGG	252
Qy	522	cttctccagttgcaaacacacccggagcattccacatraccggtacacagctatacaatccagaga	581
Db	251	CTTCTCCAGTGTGAAACACACCCGAGCAATCTCCACCATATACCGGTACACGGCTCAAAATCCAAAGA	192
Qy	582	ccatcaatcctccgtctagagttgcaaggtlbgcaaacaccccaaggtlbgctgacccaagactgca	641
Db	191	CCATCAATCCCTGCTAGATGAGTGCAGGGGTGGCAACCCCAAGGGTGGTGCACCAAGACTGCA	132
Qy	642	gaagctcctccatctttaagtgccacttaagcctcctctgacatttaactccagcatccagtg	701
Db	131	GAGTCTCTCCATCTTTCAGGTCACATTAAGCCTCTGTGCATTTAACATCAACGATCCAGTG	72
Qy	702	gtcccccagaagatccctctctctagctcctgacatgagtlbgctlgyaagaagacatccaaac	761
Db	71	GTCGCCAAGAGATCCCTTCTCCAGCCTCTGACATGAGACCGCGTGAAGAAGACATCCAAAC	12
Qy	762	aaacaagtaat	772
Db	11	AAACAAGTAAT	1

Search completed: June 3, 2001, 03:35:21
Job time: 1257 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:17:09 ; Search time 1039.85 Seconds
(without alignments)
6771.453 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806
Sequence: 1 ggaacgagggaaatctgcc.....actcaatgcagacacaaaaa 806

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 9623517 seqs, 4368049070 residues

Word size : 0

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: em_gss_fun:*
 169: em_gss_hum1:*
 170: em_gss_hum2:*
 171: em_gss_hum3:*
 172: em_gss_hum4:*
 173: em_gss_hum5:*
 174: em_gss_hum6:*
 175: em_gss_hum7:*
 176: em_gss_hum8:*
 177: em_gss_hum9:*
 178: em_gss_inv1:*
 179: em_gss_inv2:*
 180: em_gss_inv3:*
 181: em_gss_inv4:*
 182: em_gss_inv5:*
 183: em_gss_inv6:*
 184: em_gss_inv7:*
 185: em_gss_inv8:*
 186: em_gss_inv9:*
 187: em_gss_inv10:*
 188: em_gss_inv11:*
 189: em_gss_inv12:*

190: em_gss_pin1:*
 191: em_gss_pin2:*
 192: em_gss_pro:*
 193: em_gss_rod1:*
 194: em_gss_rod2:*
 195: em_gss_rod3:*
 196: em_gss_rod4:*
 197: em_gss_rod5:*
 198: em_gss_vrt1:*
 199: em_gss_vrt2:*
 200: em_gss_vrt3:*
 201: gb_gss1:*
 202: gb_gss2:*
 203: gb_gss3:*
 204: gb_gss4:*
 205: gb_gss5:*
 206: gb_gss6:*
 207: gb_gss7:*
 208: gb_gss8:*
 209: gb_gss9:*
 210: gb_gss10:*
 211: gb_gss11:*
 212: gb_gss12:*
 213: gb_gss13:*
 214: gb_gss14:*
 215: gb_gss15:*
 216: gb_gss16:*
 217: gb_gss17:*
 218: gb_gss18:*
 219: gb_gss19:*
 220: gb_gss20:*
 221: gb_gss21:*
 222: gb_gss22:*
 223: gb_gss23:*
 224: gb_gss24:*
 225: gb_gss25:*
 226: gb_gss26:*
 227: gb_gss27:*
 228: gb_gss28:*
 229: gb_gss29:*
 230: gb_gss30:*
 231: gb_gss31:*
 232: gb_gss32:*
 233: gb_gss33:*
 234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	460	57.1	460 104	AI983767 wu20a04.x
C 2	450	55.8	453 116	AW516596 xq01h07.x
C 3	429	53.2	906 141	BR899580 601682443
C 4	410	50.9	543 122	AW970357 EST82438
C 5	406	50.4	413 19	AI339648 qk63a12.x
C 6	398	49.4	450 104	AI948903 wq17c02.x
C 7	393	48.8	444 102	AI833297 at67a07.x
C 8	380	47.1	467 18	AI304380 qo59c12.x
C 9	379	47.0	430 18	AI283185 q449g09.x
C 10	377	46.8	429 102	AI813445 wj06e01.x
C 11	371	46.0	517 120	AW854263 RC3-CT025
C 12	369	45.8	420 14	AA938765 on44h11.s
C 13	366	45.4	455 102	AI833291 at68f03.x
C 14	362	44.9	467 102	AI831407 wj64a12.x
C 15	359	44.5	405 102	AI833021 at7a612.x
C 16	353	43.8	423 19	AI336470 qo61d01.x
C 17	351	43.5	466 114	AW361500 QV2-CT026
C 18	350	43.4	416 112	AW206923 UI-H-B11-

	/db_xref="taxon:9606"
	/clone="IMAGE:2748637"
	/clone_lib="Soares_NHCEC_cervical_tumor"
	/tissue_type="tumor"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: cervix; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAAGGAGCGGCCGCACAATTGTTCATTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized; constructed by Bento Soares and M.Felina Bonafide."
BASE COUNT	95 a 92 c 142 g 124 t
ORIGIN	
Query Match	55.8%; Score 450; DB 116; Length 453;
Best Local Similarity	100.0%; Pred. No. 7.7e-219;
Matches 450; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	343 ctgtctcccttcaggccttcacagcaagtgcataatgtygaagcctcatctcg 402
Dd	453 CTGTCTCCCTTTCAGCCTTCACAGCATGAGCTGCATGTGGAGGCTTATCTCGG 394
Oy	403 .ctgcaaggaccctggaaagtctccagaactcacagtccttgctcataatgtgccacaac 462
Dd	393 CTGCAAGGACCCTGGGAAGTCTCAGAACTCCACGCTGTGTCTCAATTGTCCATCAC 334
Oy	463 ttccgaagcatcatgatgaccaactcaccccaagygccctcaagtccacattgtggcc 522
Dd	333 TTTCAGAGCTAATCAAGAGCAACCTCACCCCACAGGGCTCTAGTGCACCATGTGGGGCC 274
Oy	523 tcctcagtgcaaacacacccagagatctccaccatgacgcgycfcaagctlaacaatccagaagac 582
Dd	273 TCTCCAGTGTCAAACACCGAGCATTTCCACCATGACCGGTCACAGCTACAAATCCAGAGAC 214
Oy	583 catcaatcctgttagatgtaagtcaggtgycgaagcaccacaagytgctgtgacaagaatgcyag 642
Dd	213 CATCAATCTCTGTAGATGTCAGGTGGGCAAGACCCAAAGGTGGCTGACCAAGACTGCAG 154
Oy	643 agtctcctccatcttcaggttcattcagccttcctgycalltaactaccagatccagttg 702
Dd	153 ACTGTCCTCCATCTTCAGAGTCCATTCAGCGCTCCTGGCATTTAACTAACGATCCAGTGG 94
Oy	703 tcccacaagaatcccttcctctagctcctctgcaaatgagctgtcgtggaagaagatccaaaaca 762
Dd	93 TCCCCAAGAAATCCCTTCTTAGCTCTGTACATGAGTCTGTGGAAAGAGCATCCAAACA 34
Oy	763 aaacaagtaaataataataaactcaa 792
Dd	33 AACCAATATAATAATAATAAATTAACTCNA 4
RESULT 3	
LOCUS BE899580	906 bp mRNA EST 29-SEP-2000
DEFINITION 601682443P1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:395253 5'	
ACCESSION BE899580	mRNA sequence.
VERSION BE899580.1	GI:10367234
KEYWORDS EST.	
SOURCE human.	
ORGANISM Homo sapiens	
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
TITLE NIH-MGC http://mgc.nci.nih.gov/	
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT Unpublished (1999)	
CONTACT Robert Strausberg, Ph.D.	
TEL:(301) 496-1550	

[illegible]

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 543)
Hedde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt,
I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johngett@igf.org
Plate: 276
Seq primer: Forward.
Location/Qualifiers
FEATURES
source 1..543
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE_resequences, MAGE"
/note="Vector: pBluescriptSkm"
BASE COUNT 120 a 121 c 151 g 151 t
ORIGIN
Query Match 50.9%; Score 410; DB 122; Length 543;
Best Local Similarity 100.0%; Pred. No. 2.1e-198;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 380 atgttgaggaggttattctcgcgctgcaagagccctgggaagttccagaactccagtc 439
DB 410 AGTGTGAGGAGGCTTATCTCGGGCTGCAAGACCCCTGGAAAGTTCACAACTCCACGTC 351
QY 440 ctgtctcaattgtgcataacttcagagctatcatatagccaactccaccacaagg 499
DB 350 CTTGTCTCAATGTGTCACATCACTTTCAGAGCTATCATAGGCAACCTCACCCACAGGG 291
QY 500 cctcagtcgcacacatgttggtcctctccagtcgcaaacccagagcattccaccatgacg 559
DB 290 CTTCAAGTCCCAACAGTGTGGGCTCTCCAGTCCAAACCCAGCATTCACCATGACCG 231
QY 560 gtcaaacgtcacaaatccagaagacatcaatcctgtctagagtcagagtgagcaagaccca 619
DB 230 GTCAACGCTACAAATCCAGAGACATCAATCTGCTAGAGTCAGAGGTGGCAAGCACCCA 171
QY 620 aggttgctgacaaagactgacagagctctctccatcttcaagtcattcaagctctctgc 679
DB 170 AGGGTGGCTGACCAAGACAGTCAAGAGTCTCTCATCTTAGGTCATTAGGCTCTCTGCG 111
QY 680 atttaactacagcagcatcagtggttccccaaggaatcccttcttagcctctgacatgagt 739
DB 110 ATTTAATCTACCGACATCCAGTGTGCTCCCAAGGAATCCCTTCTAGGCTCTGACATGAGT 51
QY 740 ctgctggaaagagcattccaaacaagaagtaataataataataaact 789
DB 50 CTGCTGGAAAGAGCATCCAAACAAACAAAGTAATAATAATAAATAACT 1

RESULTS 5
A1339648 413 bp mRNA EST 13-FEB-1999
LOCUS A1339648/C
DEFINITION qk63a12.x1 NCI_CGAP_C08 Homo sapiens cDNA IMAGE:1873630 3',
mRNA sequence.
ACCESSION A1339648
VERSION A1339648.1 GI:4076575
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 413)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ELN at:
www.bio.lnl.gov/bbrp/image/image.html
Insert Length: 846 Std Error: 0.00
Seq primer: -400P from Glibco
High quality sequence stop: 407.
Location/Qualifiers
FEATURES
source 1..413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1873630"
/clone_lib="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="VDH10B"
/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p773
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaudo."
BASE COUNT 82 a 82 c 128 g 121 t
ORIGIN
Query Match 50.4%; Score 406; DB 19; Length 413;
Best Local Similarity 100.0%; Pred. No. 2.3e-196;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 387 aggtcttaattctcggagctcgaagacccctgggaagttccagaactccagtcctgtct 446
DB 413 AGGCTTCAATCTCGGGCTGCAAGACCCCTGGAAAGTTCACAACTCCACGTCCTGTGT 354
QY 447 caattgtgcataacttcagagctatcatatagccaactccaccacaaggcctcagt 506
DB 353 CAATTGTGCATCACTTTCAGAGCTATCATAGGCAACCTCACCCACAGGGCTCACT 294
QY 507 ggcacacatgttggtcctctccagtcgaacacagagacattccacatgacccgtacag 566
DB 293 CGCCACCAAGTGGGCTCTCCAGTCCAAACCCAGCATTCACCATGACCGGTACAG 234
QY 567 ctcaaacctcagaagacatcaatcctgtctagagtcgaggtggaagaccccaaggtg 626
DB 233 CTACAAATCCAGAGACATCAATCTGCTAGAGTCAGAGGTGGCAAGCACCAAGGTTG 174
QY 627 ctgaccaagactgacagagctctccatcttcaagtcattcaagcctctgacattaac 686
DB 173 CTGACCAAGACTGACAGAGTCTCTCATCTTAGGTCATTAGGCTCTCTGACATTAAC 114
QY 687 taccagcatcagtggttccccaaggaatcccttcttagcctctgacatgagtctgctg 746
DB 113 TACCAAGCATCCAGTGTGCTCCCAAGGAATCCCTTCTAGGCTCTGACATGAGTGTGCG 54
QY 747 aaagagcatcacaacaagaagtaataataataataaactcaa 792
DB 53 AAAGAGCATCCAAACAAACAAAGTAATAATAATAAATAACTCA 8

[illegible]

QY	584	atcaactctgtcagagtgtagggtggacaacccccaaagggtgtgcaccagaactcgaca	643
DB	210	ATCAATCTGCTAGAGTGTACGGAGGACAACCACAAAGGTGTGCACCAACTGCCA	151
QY	644	glctcccatcttcaagtlccaattcaagtcctctgtgcatlltaactaccagcatlccagtgt	703
DB	150	GTCCTCCTCATCTTGACGTTCCATTGACGCTCCTGTGCATTTTAACATCCAGCATCCAGTGT	91
QY	704	ccccgaagaaatcccttccctagctcctctgtacatagttctgtgtgaaaagagcatccaaca	763
DB	90	CCCCAAGAATCCCTCTTAGCGCTCTCACATGAAGTCTGTGGAAAAGACATCCAACAA	31
QY	764	acaagtaaataataataaactca 792	
DB	30	ACAAATAATAATAATAATAACTCA 2	
RESULT	7		
LOCUS	A1833297/c		
DEFINITION	A1833297	EST	13-JUL-1999
ACCESSION	atc7a07.x1 Barstead colon HPLRB7	Homo sapiens	cDNA clone
VERSION	IMAGE:2377044.3,	mRNA sequence.	
KEYWORDS	A1833297		
SOURCE	A1833297.1 GI:5455277		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 444) Hillier,L., Allen,M., Bowles,L., Dubugue,T., Giesel,G., Jost,S., Kizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Merrin,M., Martin J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Thelsting,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R. WashU-NCI human EST project Unpublished (1997) Contact: Wilison RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.lml.gov) for further information. Seq primer: -40bp from Glpco.		
TITLE	JOURNAL		
COMMENT			
FEATURES	Source		
	1..444	/organism="Homo sapiens"	
	/db_xref="taxon:9606"		
	/clone_image="IMAGE:2377044"		
	/clone_lib="Barstead colon HPLRB7"		
	/sex="male"		
	/dev_stage="adult, age 25"		
	/lab_host="DH10B (phage resistant)"		
	/note="Organ: Colon; Vector: pTZ19D-Pac (Pharmacia) with a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCAGATCTCATGATGGAGCGGCGCCCTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' AATCTCATGTAT 3' and 5' ATTCTACTGT 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector. Library constructed by Bob Barstead."		
BASE COUNT	93 a	92 c	141 g 118 t
ORIGIN			
Query Match	48.8%	Score 393;	DB 102; Length 444;
Best Local Similarity	99.8%;	Pred. No. 1e-189;	
Matches 443;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
347	ctccccttcagccttcacagcagtgagctgcaatgltgtgagggcttcacctcgtgctgc	406	

```

DB 444 CTCCTCTTACACCTTCACAGCATGAGCTGAATGTTGAGGGCTTCATCTCGGGCTCC 385
    |||
QY 407 aaggacccttggaagagttccagaactccagtcctgtctcattgtgccaacttc 466
    |||
DB 384 AAGGACCCGGAAGGTTCCAGACTCCACGTCCTGTCTCAATGTTGTCATCACTTTC 325
    |||
QY 467 agagctatcatgagccaactcaccacagggcctcagtcgccaacatgtggcctc 526
    |||
DB 324 AGAGCTATCATGAGCCCAACCTCACCCACAGGGCCTCATCTCCACCAAGTGGGCTCTC 265
    |||
QY 527 cagtcacaccccgagagcttcacacatgacggtcagtcacagcacaatccagagaccac 586
    |||
DB 264 CAGTGAACACCCAGACGCTTCACCATGGCCGGTACAGCTCAATATCCAGAGACATC 205
    |||
QY 587 aatctgtcagagtgcaaggttggaagacaccccaaggtgtgctgacacagactgcagagtc 646
    |||
DB 204 AATCCTGCTAGAGTGCAAGGTGGCAAGCACCAGGGTGGCTGACCAAGACTGCAGAGTTC 145
    |||
QY 647 tcttcacatctcaggtcctcagtcctcctgagcatttaactccagacatccagtgctcc 706
    |||
DB 144 TCCTCATCTTCAGGTCATCAGCTCCTGGCATTAACTACACAGCATCCAGTGGTCCC 85
    |||
QY 707 caaggaatccctctcagtcctcagtcacatgagtcgtgtgaaagagcaccacaaacaa 766
    |||
DB 84 CAGGAGATCCCTTCAGCTCCTGACATGAGTGTGCTGGAAAGACATCCAAACAAACA 25
    |||
QY 767 agtaataataataataaactc 790
    |||
DB 24 AGTAATAATAATAATAAATTAATCT 1
    |||

```

```

RESULT 8
LOCUS AI304380/c 467 bp mRNA EST 01-FEB-1999
DEFINITION q059c12.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1912822 3',
            mRNA sequence.
ACCESSION AI304380
VERSION AI304380.1 GI:3988069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 467)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/nclogap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            CDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert length: 604 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 438.
            Location/Qualifiers
                1..467
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_image="IMAGE:1912822"
                /clone_lib="NCI_CGAP_C08"
                /tissue_type="adenocarcinoma"
                /lab_host="DH10B"
                /note="Organ: colon; Vector: p7T73D-Pac (Pharmacia) with a
                modified polylinker; 1st strand cDNA was prepared from

```

```

JOURNAL
COMMENT
AUTHORS
TITLE
FEATURES
source

```

```

BASE COUNT 96 a 97 c 148 g 126 t
ORIGIN
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo.

```

```

Query Match 47.1%; Score 380; DB 18; Length 467;
Best Local Similarity 99.8%; Pred. No. 4,5e-183;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 325 gaggtaccagcagcctcctgtctcccttcagccttcacagcagtgagtcgaatgt 384
    |||
DB 467 GAGGTACCAGCAGCCTCCTGTCCTCCCTTACAGCCTTCACAGCAGTGAAGTCATGTT 408
    |||
QY 385 gagaggttcattctcgtggttgcaagaccttggaagttccagactccagtccttgt 444
    |||
DB 407 GGAGGCTTCATCTGGGCTGCAAGACCTGGGAAAGTCCAGACTCAGCTCCTTGT 348
    |||
QY 445 cccaattgtgcatacacttcagagctatcatgagccaactcaccacaggggtccta 504
    |||
DB 347 CTCATTTGTCATCACTTTCAGAGCTATCATGAGCCAAACCTCACACCCACAGGGCTCA 288
    |||
QY 505 gtgcacacatgtggcctcctcagtgcaaacccagcagatccacatgacaggtcac 564
    |||
DB 287 GTGCGCACATGTGGGCCCTCCAGTCAAAACACGACGATTCACATGACCGGTGAC 228
    |||
QY 565 agctacaatccagagaccataatcctgtcagtgagtcagagtggaacacccaaggt 624
    |||
DB 227 ACTTAATAATCCAGACCATCATATCTCTGAGTGCAGGAGGGAACACCCAAAGGT 168
    |||
QY 625 ggctgacacagactgcagagctcctccatcttcagtcagtcacagcctctgacatta 684
    |||
DB 167 GGCTGACCAAGACTGCAAGGTCTCTCCATCTTCAGAGTCCATTCAGCCCTCCGGCATTTA 108
    |||
QY 685 actacagcagtcagtggtgcccccaaggaatcccttcagtcagtcctcagtcagtcgtgt 744
    |||
DB 107 ACTACAGCATCCAGTGGTCCCAAGAGATCCCTTCAGCTCCGACATGAGTGTCT 48
    |||
QY 745 ggaagagagcat 755
    |||
DB 47 GGAAGAGCAT 37
    |||

```

```

RESULT 9
LOCUS AI283185/c 430 bp mRNA EST 28-JAN-1999
DEFINITION qk49g09.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1872352 3',
            mRNA sequence.
ACCESSION AI283185
VERSION AI283185.1 GI:3921418
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 430)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/nclogap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            CDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
 Insert Length: 933 Std Error: 0.00
 Seq primer: -400P from Glibco
 High quality sequence stop: 403.

FEATURES

location/Qualifiers
 1. 430
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="1872352"
 /clone_lib="NCI CGAP Co8"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

88 a 90 c 133 g 119 t

ORIGIN

Query Match 47.0%; Score 379; DB 18; Length 430;
 Best Local Similarity 99.8%; Pred. No. 1.4e-182;

Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 361 ttccagcagatgagctcaatgtttgagggcttcatctcggtcgcaaggacctggaa 420
 DB 430 TTCACACAGAGAGCTCAAGTGTGAGGGCTTCATCTCGGGCTGCAAGGACCTGGGAA 371
 QY 421 agttccagaactcagctcctgtctcaattgtgcataacttcagagctatcatgag 480
 DB 370 AGTTCACAGAACTCCACTCTGTCTCAATTGTGCTCACTTCAGAGATATCATGAG 311
 QY 481 ccaacctcaaccccaagggcctcagtcgcacacatgttggcctctccagtgcaaacacc 540
 DB 310 CCAACCTCACCCCAAGGGCTCAGTCGCACATGTGGGCTCTCCAGTGCANAACACC 251
 QY 541 gagattccacatgagccgttcaacagctacaatccagaagacatcatctctagagt 600
 DB 250 GAGATTCCACCATGACCGGTGACAGCTACAAATCCAGAGCCATCATCTGTAGAGT 191
 QY 601 gcaaggtggcaagcaccgaaggtgtgtgacccaagactgcaaggtctccctccatcttcag 660
 DB 190 GCAAGGAGGCAAGCAACCAAGGTGCTGACCAAGACTGCAAGATCTCTCCATCTTCAG 131
 QY 661 gtccattcagcctctgtgcatltaactacagcatccagtggtccccaagaaatcccttc 720
 DB 130 GTCCATTACAGCTCTGTCATTTAACTACAGCATCCAGTGGTCCCAAGGAATCCCTTC 71
 QY 721 ctgacctccctggaacttggtctgctggaagaagatcccaaaaagaagtataataat 780
 DB 70 CTAACTCTCTGACTGTGAGTGTGTGGAAGAAGCATCCAAACAAAGTAATAATAAT 11
 QY 781 aaataaactc 790
 DB 10 AAATAAATCT 1

RESULT 10
 AI813445/c 429 bp mRNA EST 21-DEC-1999
 LOCUS w106601.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2402040 3',
 DEFINITION mRNA sequence.
 ACCESSION AI813445
 VERSION AI813445.1 GI:5424660
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 1 (bases 1 to 429)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL

Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
 Insert Length: 616 Std Error: 0.00
 Seq primer: -400P from Glibco
 High quality sequence stop: 389.

FEATURES

source

1. 429
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2402040"
 /clone_lib="NCI CGAP_Kid12"
 /tissue_type="2 pooled tumors (clear cell type)"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: Not I; Site: 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP_Kids was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

86 a 90 c 132 g 121 t

ORIGIN

Query Match 46.8%; Score 377; DB 102; Length 429;
 Best Local Similarity 99.8%; Pred. No. 1.5e-181;
 Matches 427; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 365 cagcagtgagactgcaatgttggagggcttcatctcggtcgcaaggacctggaaagt 424
 DB 429 CAGCAGTGAAGCTGCAATGTTGAGGGCTTCATCTCGGGCTGCAAGGACCTGGGAAAGTT 370
 QY 425 ccgaactcagctcctgtctcattgtgccatcaacttcagagctatcatgagcaa 484
 DB 369 CCAAGATCCACGCTCTGTCTCAATTGTGCCATCACTTCAGAGCTATCATAGCCAA 310
 QY 485 cctcaccacacagggcctcagtcgcacacatgttggcctctccagtgcaaacccagc 544
 DB 309 CCTCACCCCAAGGGCTCAGTCGCCACCATGTGGGCTCTCCAGTGCANAACCCAGGC 250
 QY 545 attccacatgacccggtcacaagctacaatccagaagacatcatctctgttagagtccag 604
 DB 249 ATTCCACCATGACCGGTGACAGCTACAAATCCAGAGCAATCATCTCTGAGTGCAG 190
 QY 605 gttggcaagcaccacaaaggtgtgtgacccaagactgcaaggtctccctccatcttaagtc 664
 DB 189 GAGAGCAAGACCAACCAAGGTGCTGACCAAGACTGCAAGATCTCTTCATCTTCAGTCC 130
 QY 665 attcagctctgtgcatltaactacagcatccagtggttccccaagaaatccctctag 724
 DB 129 ATTCAAGCTCTGTGCAATTTAACTACCGATCCAGTGTGCTCCCAAGGAATCCCTTCAG 70
 QY 725 ccttcctgaccttggtctgtcgtggaagaagtcctcaaaaacaagtaataataataat 784
 DB 69 CTTCTGACTAGTGTGTGTGGAAGAAGCATCCAAACAAACAGTAATAATAATAAT 10

QY	785	aaactca	792
Db	9	AAACTCAA	2
RESULT	11		
LOCUS	AM854263	517 bp	MRNA
DEFINITION	RC3-CT0254-100500-211-g12 CT0254 Homo sapiens cDNA, mRNA sequence.	EST	19-MAY-2000
ACCESSION	AM854263		
VERSION	AM854263.1	GI:7949956	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 517) Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldmann, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNLT	Proc. Natl. Acad. Sci. U.S.A.	97 (7), 3491-3496	(2000)
MEDLINE	2020263		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?r1=ct2-RC3-CT0254-100500-211-g12&t3=2000-05-10&t4=1) Seq primer: puc 18 forward High quality sequence start: 6 High quality sequence stop: 517. Location/Qualifiers 1. 517 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT0254" /dev_stage="Adult" /note="Organ: colon; Vector: puc18; Site:1; Sma1: Site:2; Sma1: A mfln1-library was made by cloning products derived from ONESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
FEATURES			
SOURCE			
BASE COUNT	106 a	118 c	172 g 121 t
ORIGIN			
Query Match	46.0%;	Score 371;	DB 120; Length 517;
Best Local Similarity	100.0%;	Pred. No. 1.8e-178;	
Matches 371;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	331	cccaacagctctctgtctcccttcagccttcacagcagctgagctgcaatgttgaggg	390
Db	434	ccacacagctctctgtctcccttcacagcagctgagctgcaatgttgaggg	375
QY	391	ctcatctcgtgagctgcaagagcctggggaagattcagaatccacatgctgtctcaat	450
Db	374	cttcatctcgtgagctgcaagagcctggggaagattcagaatccacatgctgtctcaat	315
QY	451	tgtagcatcaattctcagagctatcatgagccaactcaccaccaagagctcagtgc	510

Db	314	TGTGCATCAACTTTCAGAGCTATCATGAGCCAACTCACCACAGGAGCGTCAGTGC	255			
QY	511	accatgtggcctctccagtcgaacacacgagcatccacatgacccgtcacagctac	570			
Db	254	ACCAAGTGTGGGCTCTCCAGTGTGCAACACACCGAGCATTCACCATGACCGGTACAGCTAC	195			
QY	571	aatatcagagaccatcaatctcgtctagaagtcgtgaggtgtggaagccccaaggtgtgcga	630			
Db	194	AAATCCAGAGACCAATCAATCTCGTGAAGTGTGAGGGTGGCAAGCACCAAGGGTGGCTGA	135			
QY	631	ccaagatcgcagagctctccatcttcaggtgcatacgaacctctgtgcatctaac	690			
Db	134	CCAAGACCTGCAAGAGTCTCTCCATCTTTAGTTCATTAAGCTCCTGGCATTTAACTAC	75			
QY	691	agcatccagtg	701			
Db	74	AGCATCCAGTG	64			
RESULT	12	AA938765	420 bp	mRNA	EST	23-JUN-1998
LOCUS	AA938765/C	on44h11.91	NCI-CGAP_Co8	Homo sapiens	cDNA clone	IMAGE:1559589 3'
DEFINITION	mRNA sequence.					
ACCESSION	AA938765.1 GI:3096793					
VERSION	AA938765.1					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/hc/cgap.					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index					
JOURNAL	Unpublished (1997)					
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: www-bio.livl.gov/db/rp/image/image.html Insert Length: 613 Std Error: 0.00 Seq primer: -40m13 fwd. E7 from Amersham High quality sequence stop: 403.					
FEATURES	Location/Qualifiers					
Source	1..420					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="IMAGE:1559589"					
	/clone_1id="NCI CGAP_Co8"					
	/tissue.type="adenocarcinoma"					
	/lab_host="DH10p"					
	/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."					
BASE COUNT	86 a	88 c	130 g	116 t		
ORIGIN						
Query Match	45.8%	Score 369,	DB 14,	Length 420;		
Best Local Similarity	99.8%;	Pred. No. 1.9e-177;				

	Matches	419;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Oy	371	ttagctgtcaaatgttggagggtcattatcttcggtctgcgaagacccttggaagtccaga	430							
Dd	420	ttagctgtcaaatTTTGAAGGGCTTCATCTCGGCTCCAAAGACCTTGGAATTCCAGNA	361							
Oy	431	cfcacagctccttgttcataattgtgccatcaaatcttcagaagatatcatgtgccaacctac	490							
Dd	360	CTCCACGTCCTGTGTCAATTGTGCCATCACAATTCTTCAGAGCATATCATAGCCAACCTCAC	301							
Oy	491	cccacaagggtcctcagtcgcccaccaaactgtgtggctctccasgtgtcaaacaccagagattcca	550							
Dd	300	CCCACAGGGCCCTCAATGCACCACCATGTGGGCTCTCCACGTGCMAAACCCACCGACATTCCA	241							
Oy	551	ccatcacccgttcaacagctcaacaatccagagaccatcaatctccgtcagaagtgcgaaggtgc	610							
Dd	240	CCATACCCGGTCCACAGCTCATCAAAATCCAGAAGACATCAATCTCTGTAGTAGTCAGGGAAGGC	181							
Oy	611	aagcaaccaaggtgtgctgcacccaagctcagagctcctccatcttaagtlccattcag	670							
Dd	180	AAGCACCCCAAGGGTGGCTGACCAACAATGACAGACTCTCCTCATCTTCAGGTCATTGAG	121							
Oy	671	ccctctgagcatttaactctccagcatcccaagtgtgtccccaaagsaatcccttcagctcctc	730							
Dd	120	CCTCTGGCATTTAACTACACACATCCAGCTCAGGTGCTCCCAAAGGAAATCCCTTCAGGCTCCT	61							
Oy	731	gacatgagctctgtctggaagaagcatccacaacaacaagaattaataataataaataaac	790							
Dd	60	GACATGAGCTCTGTCTGGAAGAAGACATCCAAACAACAATAAATAATAATAATAATTAATC	1							
RESULT 13										
LOCUS	AI832391/c									
DEFINITION	at68f03.x1 Birstead colon HPRB7 Homo sapiens CDNA clone									
VERSION	AI832391									
KEYWORDS	AI832391.1 GI:5454380									
SOURCE	EST.									
ORGANISM	Homo sapiens									
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.									
COMMENT	1 (bases 1 to 455) Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacey,M., Le,N., Lennon,G., Marra,M., Martin White,Y., Wylie,T., Waterston,R. and Wilson,R. WashU-NCI human EST project unpublished (1997) Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu This clone is available royalty-free through LINT ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40UP from GIDCO High quality sequence stop: 420. Location/Qualifiers 1..455 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="IMAGE:2377181" /clone_id="Birstead colon HPRB7" /sex="male" /dev_stage="adult, age 25" /lab_host="DH10B (phage resistant)." /note="Organ: colon; Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCACTCACTGAAGTGAAGGAGCGGCCCTTTTTTTTTTTTTTTTTTTTTT									

BASE COUNT	97 a	94 c	145 g	119 t	
ORIGIN					
Query Match		45.4%	Score 366;	DB 102;	Length 455;
Best Local Similarity		99.8%;	Prod. No. 6.4e-176;		
Matches 416;	Conservative	0;	Mismatches 1;	Indels	0;
					Gaps 0;
3']					
double-stranded cDNA was ligated to Eco RI adaptors					
15' ATTACTACTACTACT 3' and 5' ATTACTAGTG 3']					
digested with Not I and cloned into the Not I and Eco RI sites of					
the modified pvr713 vector. Library constructed by Bob					
Barsstead."					
0y	341	tcctctccccccttcacagcttcacagcagagtgatgtttgagggtctatccg	400		
Db	455	TCCTCTCTCCCTTCACCTTCACAGCAGTAGTGCAATTTGAGGGCTTCATCTCG	396		
Qy	401	ggctcgaagagacccttggaaagtctcagaactccacgtctctgtctcaattgtgcatca	460		
Db	395	GGCTGCAAGAGACCCTGGGAAAGTTCCAAACATCCCGCTCTGTCTCAATTTGGCATCA	336		
Qy	461	acttcagaagctatcatatagccaacctacccccaagggtcctcagtcgccaacatgtgg	520		
Db	335	ACTTTCAAGAGCTATCATATGAGCCAACTCACCCCAAGGAGGCTCACTGCGCAACATGTGG	276		
Qy	521	cctctccagtgctcaaacaccagcagatcttcacacatgacggtctacagctacaaatccag	580		
Db	275	CCTCTCCAGTCCAAACCCAGACATTTCCACCATATACCGGTACACATCAATTCACAG	216		
Qy	581	accatcaatctgtcagaggttcagaggttggcaagaccccaagggtgtgtcgaaccaagactgc	640		
Db	215	ACCATCAATCTCTCTAGAGTGTGAGGGTGGCAAGCACCCAAAGGTGGCTGACCAAGACTGC	156		
Qy	641	agaagctctccatcttcagagtcacatlcacagctcctctgtgcatltaactacacagcatccagt	700		
Db	155	AGAGTCTCTCTCATCTTTCAGGTGCAGGGTGGCAAGCACCCAAAGGTGGCTGACCAAGACTGC	96		
Qy	701	ggtccccaaggaatccctcttcctagctcctctgaacatgagtcgtctggaaagagcatcc	757		
Db	95	GGTCCCAAGGAATCCCTTCAGCTCTCTGACATGAGTCTGTGGAAAGACATCC	39		
RESULT 14					
AI831407/c					
LOCUS	AI831407	467 bp	mRNA	EST	21-DEC-1999
DEFINITION	w164a12.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407582 3', mRNA sequence.				
ACCESSION	AI831407				
VERSION	AI831407.1	GI:5452078			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 467)				
AUTHORS	NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LuNL at: www-bio.llnl.gov/bdrip/image/image.html Insert length: 756 Std Error: 0.00 Seq primer: -400P from Glibco High quality sequence stop: 458.				

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 04:13:45 ; Search time 48.53 Seconds
(without alignments)
195.628 Million cell updates/sec

Title: US-09-599-087-5

Perfect score: 442
Sequence: 1 MRLVLSLCLILLCFSIF.....PCKLEPRRLVVPALPGV 81

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_15:*
2: SP_Archaea:*
3: SP_Bacteria:*
4: SP_Fungi:*
5: SP_Human:*
6: SP_Invertebrate:*
7: SP_Mammal:*
8: SP_MNC:*
9: SP_Organelle:*
10: SP_Phage:*
11: SP_Plant:*
12: SP_Rodent:*
13: SP_Unclassified:*
14: SP_Vertebrate:*
15: SP_Virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	17.0	763	10 Q9LW12	Q9LW12 Oryza sativ
2	75	17.0	763	10 Q9LW12	Q9LW12 Oryza sativ
3	75	17.0	763	10 Q9LW12	Q9LW12 Oryza sativ
4	71.5	16.2	714	11 P70593	P70593 ratius norv
5	70.5	16.0	523	4 Q00480	Q00480 homo sapien
6	70	15.8	86	6 Q9N0D5	Q9N0D5 macaca fasc
7	69.5	15.7	524	10 Q9LW12	Q9LW12 Oryza sativ
8	67	15.2	864	10 Q9LW12	Q9LW12 Oryza sativ
9	66	14.9	1058	5 Q9VF12	Q9VF12 dirosophila
10	65.5	14.8	509	10 Q23391	Q23391 arabidopsis
11	65	14.7	948	4 Q9Y5H9	Q9Y5H9 homo sapien
12	65	14.7	948	4 Q9Y5H9	Q9Y5H9 homo sapien
13	64.5	14.6	513	10 Q9LW12	Q9LW12 Oryza sativ
14	62.5	14.1	417	4 Q9N0D5	Q9N0D5 macaca fasc
15	62	14.0	586	3 Q9P985	Q9P985 saccharomyc
16	62	14.0	586	3 Q9P985	Q9P985 saccharomyc
17	62	14.0	586	3 Q9P985	Q9P985 saccharomyc
18	61.5	13.9	143	4 Q9N0D5	Q9N0D5 macaca fasc
19	61.5	13.9	146	11 Q63552	Q63552 ratius norv

20	61.5	13.9	211	2 Q50823	Q50823 borrelia bu
21	61.5	13.9	394	10 Q9LW12	Q9LW12 Oryza sativ
22	61.5	13.9	527	4 Q00475	Q00475 homo sapien
23	61.5	13.9	529	4 P78408	P78408 homo sapien
24	61	13.8	154	11 Q63593	Q63593 ratius norv
25	61	13.8	165	10 Q9ZQ38	Q9ZQ38 arabidopsis
26	61	13.8	622	13 Q9LW12	Q9LW12 Oryza sativ
27	60.5	13.7	225	4 Q15358	Q15358 torpeda mar
28	60.5	13.7	868	11 Q9N0D5	Q9N0D5 macaca fasc
29	60.5	13.7	868	11 Q9N0D5	Q9N0D5 macaca fasc
30	60	13.6	287	10 Q9SR31	Q9SR31 arabidopsis
31	60	13.6	477	10 Q9M1K5	Q9M1K5 arabidopsis
32	60	13.6	513	10 Q9XHC6	Q9XHC6 glycine max
33	60	13.6	615	2 P94349	P94349 bacillus st
34	60	13.6	842	4 Q15053	Q15053 homo sapien
35	60	13.6	950	4 Q9Y5H9	Q9Y5H9 homo sapien
36	60	13.6	1055	10 Q9SVF0	Q9SVF0 aeropyrum p
37	59.5	13.5	117	1 Q9YBL9	Q9YBL9 aeropyrum p
38	59.5	13.5	394	2 Q9ZAM0	Q9ZAM0 sphingomona
39	59.5	13.5	445	4 Q9NP18	Q9NP18 homo sapien
40	59.5	13.5	461	2 Q9KSJ1	Q9KSJ1 vibrio chol
41	59.5	13.5	657	4 Q14226	Q14226 homo sapien
42	59.5	13.5	878	4 Q9UFU7	Q9UFU7 homo sapien
43	59.5	13.5	893	5 Q9N3E6	Q9N3E6 caenorhabdi
44	59.5	13.5	1172	4 Q14947	Q14947 homo sapien
45	59	13.3	113	8 Q9T3F9	Q9T3F9 nephrolepim

ALIGNMENTS

RESULT 1
Q9LW12 PRELIMINARY; PRT: 763 AA.
AC Q9LW12:
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-OCT-2000 (TREMREL. 15, Last annotation update)
DE SIMILAR TO ANTIRRHINUM MAJUS TRANSPOSON TAM3 GENE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
RT clone: P0675A05."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002071; BAA95877.1; -
SQ SEQUENCE 763 AA; 84207 MW; 0A8C161720B80636 CRC64;

Query Match 17.0%; Score 75; DB 10; Length 763;

Best Local Similarity 31.2%; Pred. No. 0.57;
Matches 24; Conservative 8; Mismatches 19; Indels 26; Gaps 5;

QY 22 TEGKRPRKAW-----SGRTRL--C--CHRVSPSTNLKGH---HVLCKPC 63
DB 68 TNKKTKTSKVDDEFELEYETTNENRVSACKNCHKTLARSAGTGLHRLHNSKPR 127
QY 64 KLEPEPRLMVVPALPG 80
DB 128 KLG-----SNALPG 136

RESULT 2
Q9LW12 PRELIMINARY; PRT: 763 AA.
ID Q9LW12:
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)

RESULT	4	
P70593		
ID	P70593	PRELIMINARY;
AC	P70593;	PRT;
DT	01-FEB-1997	714 AA.
	(TIREMBLrel. 02, Created)	
	(TIREMBLrel. 02, Last sequence update)	

	RESULT	6
Q9NOD5		
ID	Q9NOD5	PRELIMINARY; PRT: 86 AA.
AC	Q9NOD5:	
DT	01-OCT-2000 (Tremblrel. 15 Created)	
PT	01-OCT-2000 (Tremblrel. 15 Last sequence update)	
DI	01-OCT-2000 (Tremblrel. 15, Last annotation update)	
DE	UNNAED PROTEIN PRODUCT.	
OS	Macaqa fascicularis (Crib eating macaque) (Cynomolgus monkey).	
OC	Eukaryota; Metazoa; Chordata; Crinata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;	

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003711; AAF5250.1.
 DR FLYBASE: FBgn0038383; CG5552.
 SQ SEQUENCE 1058 AA; 118060 MW; BBD95D5819753BE9 CRC64;

 Query Match 14.9%; Score 66; DB 5; Length 1058;
 Best Local Similarity 43.8%; Pred. No. 9.2;
 Matches 14; Conservative 3; Mismatches 13; Indels 2; Gaps 1;

 QY 42 HRVSPNSTNL-KGHWRLCKPCKLEPPRL 71
 DB 859 HRTSSSTTIAEPHHPGCHPCKTDPLPWL 890

 RESULT 10
 ID 023391 PRELIMINARY; PRT; 509 AA.
 AC 023391;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE CYTOCHROME P450 LIKE PROTEIN.
 GN DL3720W OR AT4G15350.
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 CC Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;
 RX SEQUENCE FROM N.A.
 RA Bevan M., Stekema W., Murphy G., Mambutt R., Pohl T., Terry N.,
 RA Kreis M., Kavanagh T., Enlian K.D., Rieger M., James R.,
 RA Poudoumench P., Hatzopoulos P., Obermaier B., Duesterhoft A.,
 RA Jones J., Palme K., Ansong W., Delsen Y.M., Bancroft I., Mewes H.W.,
 RA Schueller C., Chaiwatzis N.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z97338; CAB4398.1;
 DR EMBL: AL161541; CAB78577.1;
 DR INTERPRO: IPR001128;
 DR INTERPRO: IPR002401;
 DR INTERPRO: IPR002403;
 DR PFM: PF00067; P450.1.
 DR PRINTS: PR00385; P450.
 DR PRINTS: PR00463; EP450I.
 DR PRINTS: PR00465; EP450IV.
 SQ SEQUENCE 509 AA; 58395 MW; CA0D86CE254EE76B CRC64;

 Query Match 14.8%; Score 65.5; DB 10; Length 509;
 Best Local Similarity 30.6%; Pred. No. 5.6;
 Matches 19; Conservative 9; Mismatches 11; Indels 23; Gaps 3;

 QY 1 MLVLVSLLCIL-LTCSIF-----STEGRRPAKMSGRTRLCCHRVSPSTNLK 53
 DB 1 MAVLIIFLLCLSLTSLCTSLFPMKXDSRDGRLP-----PSPSLPII 44

 QY 54 GH 55
 DB 45 GH 46

 RESULT 11

ID 075287 PRELIMINARY; PRT; 824 AA.
 AC 075287;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE KIA0345-LIKE 12.
 GN PCDH-ALPHA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RA Kimmey W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
 RA Kadner K., Miguel T., Miller C., Piltuck S., Pollard M., Rojesh H.,
 RA Subramanian S., Martin C.H.;
 RT "Sequencing of human chromosome 5."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rieke D.O.;
 RT "Large Scale Sequence Analysis and Annotation with the Sequence
 Comparison Analysis (SCAN) system."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99308636; PubMed-10380929;
 RA Wu Q., Maniatis T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AC005609; AC034324.1;
 DR EMBL: AF152480; AAD3741.1;
 DR INTERPRO: IPR002126;
 DR PFM: PF00028; cadherin; 5.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 824 AA; 88866 MW; DF8CEA0DE1ACA25D CRC64;

 Query Match 14.7%; Score 65; DB 4; Length 824;
 Best Local Similarity 39.3%; Pred. No. 9.8;
 Matches 22; Conservative 4; Mismatches 16; Indels 14; Gaps 3;

 QY 3 LVLVSLLCILLCFSLFSTEGRRPAK-----AMS---GRTRLCHRPV 45
 DB 710 LVLVLLVLYTALRC-SVPTGAPAPGKPTLVCSAVGSMYSQGRQRCVSGEDP 764

 RESULT 12
 ID 09Y5H9 PRELIMINARY; PRT; 948 AA.
 AC 09Y5H9;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE PROTOCADHERIN ALPHA 2.
 GN PCDH-ALPHA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RA Tissue-BRAIN;
 RX MEDLINE-99308636; PubMed-10380929;
 RA Wu Q., Maniatis T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF152310; AAD43704.1; -
 DR INTERPRO: IPR002126; -
 DR PFMW: PF00028; cadherin.5.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN.5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 KW SEQUENCE 948 AA; 102062 MW; 8D2E9644982AE59E CRC64;

Query Match 14.7%; Score 65; DB 4; Length 948;
 Best Local Similarity 39.3%; Pred. No. 11;
 Matches 22; Conservative 4; Mismatches 16; Indels 14; Gaps 3;

OY 3 ILVLSILCLILLCFSIFSTEGKRRPAK-----AMS---GRTFLCCHRP 45
 DB 710 LVLVLYLTALRC-SVPTTEGARAPGKPTLVCSAVGSMSYSQGRRCVCSGEDP 764

RESULT 13

O9LTY8 PRELIMINARY; PRT: 513 AA.

AC O9LTY8: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CYTOCHROME P450-LIKE PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-COLUMBIA.
 RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RT [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN-COLUMBIA.
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
 RT TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AP000383; BAB01870.1; -
 SO SEQUENCE 513 AA; 58260 MW; F295D1F35BD3142 CRC64;

Query Match 14.6%; Score 64.5; DB 10; Length 513;
 Best Local Similarity 39.6%; Pred. No. 7.4;
 Matches 19; Conservative 6; Mismatches 14; Indels 9; Gaps 3;

OY 9 LILCL-ILCFSTFSTEGKRRPAKMSGRRLCCHRPSPSTNLKGH 55
 DB 14 LILCFSLVCYSLFP---RRP---SSRDAHGCDLPSPSPVLIGH 53

RESULT 14

O9NMJ8 PRELIMINARY; PRT: 417 AA.

AC O9NMJ8: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CDNA FLJ20798 F1S, CLONE ADS02031.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE-ADIPOSE TISSUE;
 SO

RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
 RA Hirao M., Ohmori Y., Ota T., Suzuki Y., Odayashi M., Nishi T.,
 RA Shibahara T., Tanaka T., Nakamura Y., Isegaki T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK000805; BAA91382.1; -
 KW SEQUENCE 417 AA; 43432 MW; 85AD6BC3F7E66E8 CRC64;

Query Match 14.1%; Score 62.5; DB 4; Length 417;
 Best Local Similarity 32.7%; Pred. No. 11;
 Matches 18; Conservative 8; Mismatches 24; Indels 5; Gaps 2;

OY 28 PAKMSGRRTFLCCHRPSPSTNLKGHHVRLCKPKLT--EPEPLVYVGPALPQ 80
 DB 123 PAKMAKAKKQELQHANSP---LLRGSLSLRLACELNRSKSLNSQSPAVPQ 174

RESULT 15

O9P986 PRELIMINARY; PRT: 586 AA.

AC O9P986: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE ENDONULEASE FOR MATING-TYPE CONVERSION.
 GN SC-HO.
 OS Saccharomyces pastorianus (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=27292;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-KBY001;
 RA Tamai Y., Kanai K., Umemoto N., Tomizuka K., Kaneo Y.;
 RT "Diversity of the HO gene encoding an endonuclease for mating-type
 RT conversion in the bottom fermenting yeast Saccharomyces pastorianus."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB027449; BAA94595.1; -
 SO SEQUENCE 586 AA; 66087 MW; B7D01823E1AE9DE2 CRC64;

Query Match 14.0%; Score 62; DB 3; Length 586;
 Best Local Similarity 31.7%; Pred. No. 17;
 Matches 20; Conservative 4; Mismatches 17; Indels 22; Gaps 5;

OY 30 KANSGRRTFLC--CHRVSPSPSTNLKGHHVRLCKPKCLEPFR-----LMVYVPG 76
 DB 499 KDMNG-KNRVCARCY-----GRYKFSGH--CINCKYVEARVRKRAKDKGKLGITPE 549

OY 77 ALP 79
 DB 550 GLP 552

Search completed: June 3, 2001, 04:37:23
 Job time: 1418 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 04:34:40 ; Search time 30.93 Seconds

(without alignments)
89.709 Million cell updates/sec

Title: US-09-599-087-5

Perfect score: 442

Sequence: 1 MRLVLVSLICILLCEFSIF.....PKLEPPRLMTVPGALPGV 81

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	15.6	865	1	CNSA_BOVIN
2	66.5	15.0	462	1	US29_HCMVA
3	62.5	14.1	146	1	SMRL_RAT
4	62	14.0	586	1	HO_YEAST
5	62	14.0	632	1	NTG3_HUMAN
6	60.5	13.7	247	1	MOG_HUMAN
7	60.5	13.7	470	1	NOS2_ONCMY
8	60.5	13.7	782	1	SM4B_MOUSE
9	60.5	13.7	862	1	CD22_MOUSE
10	60	13.6	477	1	URT2_DESKO
11	59.5	13.5	1172	1	TSP2_MOUSE
12	59.5	13.5	1210	1	EGFR_HUMAN
13	59	13.3	55	1	ATP8_HUMAN
14	59	13.3	365	1	EXH1_HUMAN
15	59	13.3	391	1	GAT5_CHICK
16	59	13.3	477	1	URT1_DESKO
17	58.5	13.2	130	1	YKD6_YEAST
18	58.5	13.2	616	1	MUTA_STROM
19	58.5	13.2	1173	1	TSP1_XENLA
20	58	13.1	248	1	UL24_EBV
21	58	13.1	364	1	YHIM_ECOLI
22	58	13.1	514	1	MPA2_CRYDA
23	57.5	13.0	88	1	TOXK_WILMR
24	57.5	13.0	516	1	GIC5_SOYBN
25	57.5	13.0	559	1	TPA_RAT
26	57.5	13.0	889	1	RPAL_METVA
27	57	12.9	763	1	EXM1_HUMAN
28	57	12.9	841	1	IE63_MCWVS
29	57	12.9	993	1	VIA_TAV
30	56.5	12.8	566	1	EPOR_MOUSE
31	56.5	12.8	507	1	TPA_BOVIN
32	56.5	12.8	706	1	SM2A_DROME
33	56.5	12.8	1338	1	VGRL_HUMAN

ALIGNMENTS

RESULT	ID	CNSA_BOVIN	STANDARD:	PRT:	865 AA.
AC	Q28156:				
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	CGMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17) (CGB-PDE)				
DE	(CGMP-BINDING CGMP-SPECIFIC PHOSPHODIESTERASE).				
GN	PDE5A OR PDE5.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OX	NCBI_TaxID=9913;				
RM	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Lung;				
RX	MEDLINE=94043054; PubMed=8226796;				
RA	McAllister-Lucas L.M., Sonnenburg W.K., Kadlecsek A., Seger D.,				
RA	Trong H.L., Colbran J.L., Thomas M.K., Walsh K.A., Francis S.H.,				
RA	Corbin J.D., Beavo J.A.;				
RT	"The structure of a bovine lung CGMP-binding, CGMP-specific phosphodiesterase deduced from a cDNA clone";				
RL	J. Biol. Chem. 268:22863-22873(1993).				
RN	[2]				
RP	METAL-BINDING; PubMed=8077192;				
RX	MEDLINE=94357882; PubMed=8077192;				
RA	Francis S.H., Colbran J.L., McAllister-Lucas L.M., Corbin J.D.;				
RT	"Zinc interactions and conserved motifs of the CGMP-binding CGMP-specific phosphodiesterase suggest that it is a zinc hydrolase.";				
RL	J. Biol. Chem. 269:22477-22480(1994).				
RN	[3]				
RP	MUTAGENESIS.				
RX	MEDLINE=96107229; PubMed=8530505;				
RA	McAllister-Lucas L.M., Halk T.L., Colbran J.L., Sonnenburg W.K.,				
RA	Seger D., Turko I.V., Beavo J.A., Francis S.H., Corbin J.D.;				
RT	"An essential aspartic acid at each of two allosteric CGMP-binding sites of a CGMP-specific phosphodiesterase.";				
RL	J. Biol. Chem. 270:30671-30679(1995).				
RN	[4]				
RP	MUTAGENESIS.				
RX	MEDLINE=96355629; PubMed=8703039;				
RA	Turko I.V., Halk T.L., McAllister-Lucas L.M., Burns F., Francis S.H.,				
RA	Francis S.H., Corbin J.D.;				
RT	"Identification of key amino acids in a conserved CGMP-binding site of CGMP-binding phosphodiesterases. A putative NXXD motif for CGMP binding.";				
RL	J. Biol. Chem. 271:22240-22244(1996).				
RN	[5]				
RP	PHOSPHORYLATION; AND MUTAGENESIS.				
RX	MEDLINE=98109724; PubMed=9445376;				
RA	Turko I.V., Francis S.H., Corbin J.D.;				
RT	"Binding of CGMP to both allosteric sites of CGMP-binding CGMP-specific phosphodiesterase (PDE5) is required for its phosphorylation.";				

RL Biochem. J. 329:505-510(1998).
CC -1- FUNCTION: PLAYS A ROLE IN SIGNAL TRANSDUCTION BY REGULATING THE
CC INTRACELLULAR CONCENTRATION OF CYCLIC NUCLEOTIDES. THIS
CC PHOSPHODIESTERASE CATALYZES THE SPECIFIC HYDROLYSIS OF CGMP TO 5'-
CC GMP.
CC -1- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O -
CC GUANOSINE 5'-PHOSPHATE.
CC -1- COFACTOR: REQUIRES DIVALENT CATIONS. ZINC IONS ARE REQUIRED FOR
CC MAXIMUM ACTIVITY. MANGANESE, MAGNESIUM AND COBALT ALSO SUPPORT
CC CATALYSIS BUT AT MUCH HIGHER CONCENTRATIONS.
CC -1- ENZYME REGULATION: MOST POTENTIALLY INHIBITED BY ZAPRINAST AND
CC DIPYRIDAMOLE.
CC -1- PATHWAY: CYCLIC NUCLEOTIDE METABOLISM.
CC -1- DOMAIN: COMPOSED OF A C-TERMINAL CATALYTIC DOMAIN CONTAINING TWO
CC POTENTIAL DIVALENT METAL SITES AND AN N-TERMINAL REGULATORY DOMAIN
CC WHICH CONTAINS TWO HOMOLOGOUS ALLOSTERIC CGMP-BINDING REGIONS, A
CC AND B.
CC -1- PTM: PHOSPHORYLATION IS REGULATED BY BINDING OF CGMP TO THE TWO
CC ALLOSTERIC SITES.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L16545; AAB0990.1; -
DR InterPro: IPR002073; -
DR InterPro: IPR003018; -
DR Pfam: PF01590; GAF; 2.
DR Pfam: PF00233; PDSEase; 1.
DR PRINTS: PR00387; PDSETERASE1.
DR PROSITE: PS00126; PDSEASE1; 1.
KW Hydrolyase; CGMP; CGMP-binding; Phosphorylation; Zinc; Repeat.
FT MOI_RES 92 92
FT METAL 603 603 ZINC 1 (POTENTIAL).
FT METAL 607 607 ZINC 1 (POTENTIAL).
FT METAL 632 632 ZINC 1 (POTENTIAL).
FT METAL 643 643 ZINC 2 (POTENTIAL).
FT METAL 647 647 ZINC 2 (POTENTIAL).
FT METAL 672 672 ZINC 2 (POTENTIAL).
FT METAL 672 672 ZINC 2 (POTENTIAL).
FT NP_BIND 228 311 CGMP.
FT NP_BIND 410 500 CGMP.
FT BINDING 276 276 CGMP.
FT BINDING 277 277 CGMP.
FT BINDING 289 289 CGMP.
FT BINDING 478 478 CGMP.
FT DOMAIN 578 843 CATALYTIC (BY SIMILARITY).
FT MUTAGEN 276 276 N->A: DECREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 277 277 K->A: DECREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 277 277 K->R: SLIGHT INCREASE IN CGMP-BINDING.
FT MUTAGEN 289 289 D->A: DECREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 289 289 D->N: INCREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 290 290 E->A: NO CHANGE IN CGMP-BINDING.
FT MUTAGEN 478 478 D->A: INCREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY. PHOSPHORYLATED AT
FT LOWER CONCENTRATIONS OF CGMP.
SQ SEQUENCE 865 AA; 98626 MW; 2EF744B2990B4F7 CRC64;

Query Match 15.68; Score 69; DB 1; Length 865;
Best Local Similarity 36.18; Pred. No. 3;
Matches 22; Conservative 4; Mismatches 27; Indels 8; Gaps 5;

QY 17 FSISTEGRKRRPAKWSRTRTLCCHRVPSNINIKGHHVRLCKPCKLEPEPRL-WVVP 75

DB 32 FSYFVRKCTREVMANFAERY-----HTIPVCKE-GIKS-HIESCS-CLOSPRESSVP 84
QY 76 G 76
DB 85 G 85

RESULT 2
US29_HCMVA STANDARD; PRT; 462 AA.
ID US29_HCMVA
AC P09705;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 01-FEB-1991 (Rel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN HHRF4.
GN US29.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_Taxid:10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:87169717; PubMed:3031311.
RT Weston K., Barrell B.G.;
RT "Sequence of the short unique region, short repeats, and part of the
RT long repeats of human cytomegalovirus";
RT J. Mol. Biol. 192:177-208(1986).
RN [2]
RP COMPLETE GENOME.
RX MEDLINE:90269039; PubMed:2161319;
RA Chee M.S., Bankier A.T., Beck S., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169";
RT Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17403; CA35261.1; -
DR EMBL: X04650; CA28339.1; -
DR PIR: D27216; OQBED4.
DR PIR: S09943; S09943.
KW Hypothetical protein.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 462 AA; 51066 MW; D8D070B42EB6B289 CRC64;

Query Match 15.08; Score 66.5; DB 1; Length 462;
Best Local Similarity 25.88; Pred. No. 3.2;
Matches 23; Conservative 7; Mismatches 42; Indels 17; Gaps 3;

QY 4 LVSSLLCILLIC--FSISTEGRKRRPAKWS-----GRTRLCCHR--VPS 46
DB 258 LCVDLVLCVLLALLLELVMEAVRHPLFWRRVALSPSTKYDRAVKLCLERRMFGILP 317
QY 47 PNSTNLKGHHVRLCKPCKLEPEPRLWVVP 75
DB 318 PPSVAPPEKEKELEPAQAALSPPLTTWSLP 346

RESULT 3
SMRL_RAT STANDARD; PRT; 146 AA.
ID SMRL_RAT

AC P13432;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SMRI PROTEIN PRECURSOR (VCS-ALPHA 1).
 GN VCSA1 OR SMRI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Submaxillary gland;
 RX MEDLINE=89042220; PubMed=3166744;
 RA Rosinski-Chupin I., Tronik D., Rougeon F.;
 RT "High level of accumulation of a mRNA coding for a precursor-like
 RT protein in the submaxillary gland of male rats.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:8553-8557(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9103875; PubMed=2125424;
 RA Rosinski-Chupin I., Rougeon F.;
 RT "The gene encoding SMRI, a precursor-like polypeptide of the male rat
 RT submaxillary gland, has the same organization as the
 RT preprothymotropin-releasing hormone gene.";
 RL DNA Cell Biol. 9:553-559(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NISTAR;
 RX MEDLINE=96032360; PubMed=7557446;
 RA Courty Y., Rosinski-Chupin I., Rougeon F.;
 RT "Various transcripts are generated from the VCSA1 gene by alternative
 RT splicing and poly(A) processing in the rat submandibular gland.";
 RL Gene 162:291-296(1995).
 RN [4]
 RP SEQUENCE OF 23-33.
 RC TISSUE-Submandibular gland;
 RX MEDLINE=9415861; PubMed=8112327;
 RA Rougeon C., Rosinski-Chupin I., Njamkepo E., Rougeon F.;
 RT "Selective processing of submandibular rat 1 protein at dibasic
 RT cleavage sites. Salivary and bloodstream secretion products.";
 RL Eur. J. Biochem. 219:765-773(1994).
 CC -1- FUNCTION: UNKNOWN. MALE-SPECIFIC FUNCTION, BUT THE PRESENCE OF A
 CC TETRAPEPTIDE MOTIF SURROUNDED BY EXPOSED PAIRED BASIC RESIDUES
 CC SUGGESTS THAT IT COULD BE PROCESSED, AND CONSEQUENTLY ITS
 CC MATURATION PRODUCTS MAY HAVE A PHYSIOLOGICAL ROLE IN MALE RATS.
 CC -1- INDUCTION: HIGH LEVEL OF INDUCTION BY ANDROGENS.
 CC -1- P-TM: SEVERAL O-LINKED GLYCOSYLATION SITES MIGHT BE PRESENT IN THE
 CC C-TERMINUS OF SMRI.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M63112; AAA2153.1; -
 CC EMBL: M59467; AAA2154.1; -
 CC EMBL: X84997; CAA5935.1; -
 CC EMBL: X52467; CAA36705.1; -
 CC EMBL: A07543; CAA00668.1; -
 CC PIR: A31347; A31347.
 KM Submandibular gland; Signal: Glycoprotein;
 KW Cleavage on pair of basic residues; Polymorphism.
 FT SIGNAL 1 22
 FT CHAIN 23 146
 FT PEPTIDE 23 33 SMRI PROTEIN.
 FT PEPTIDE 29 33 SMRI-RELATED UNDECAPEPTIDE.
 FT CARBOHYD 129 129 SMRI-RELATED PENTAPEPTIDE.
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 114 114 H -> L.

SO SEQUENCE 146 AA; 15970 MW; F31B619A3BD85B5C CRC64;
 Query Match 14.1%; Score 62.5; DB 1; Length 146;
 Best Local Similarity 34.9%; Pred. No. 3.2;
 Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6;
 OY 1 MRLIVSSLLCILLCFISFTEGKRPAKWSGRTR--LCGRV--PSPNSTNKG 54
 DB 1 MSLYLIFGLMILLACFO--SGEGVRGPRQHNRRQDPSTLPHYLGLQDPNGQT-- 56
 OY 55 HNRVLCCKPKCKEPRRLMV--VPG 76
 DB 57 GVTITPLNLP--PRVLVNLPG 77
 RESULT 4
 HO_YEAST STANDARD; PRT; 586 AA.
 ID HO_YEAST
 AC P0932; Q12183;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HOMOTHALLIC SWITCHING ENDONUCLEASE.
 GN HO OR YDL227C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87089786; PubMed=3025649;
 RA Russell D.W., Jensen R., Zoller M.J., Burke J., Errede B., Smith M.,
 RA Herskowitz I.;
 RT "Structure of the Saccharomyces cerevisiae HO gene and analysis of
 RT its upstream regulatory region.";
 RL Mol. Cell. Biol. 6:4281-4294(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=96120866; PubMed=8590483;
 RA Meiron H., Nahon E., Raven D.;
 RT "Identification of the heterothallic mutation in HO-endonuclease of
 RT S. cerevisiae using HO/ho chimeric genes.";
 RL Curr. Genet. 28:367-373(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Rasmussen S.W.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP REVIEW.
 RX MEDLINE=91171884; PubMed=2005783;
 RA Herskowitz I., Jensen R.;
 RT "Putting the HO gene to work: practical uses for mating-type
 RT switching.";
 RL Meth. Enzymol. 194:132-146(1991).
 CC -1- FUNCTION: INITIATION OF MATING TYPE INTERCONVERSION. THIS PROTEIN
 CC IS A SITE-SPECIFIC ENDONUCLEASE THAT CLEAVES A SITE IN THE MAT
 CC LOCUS ON CHROMOSOME III. THE DOUBLE-STRAND BREAK IS FOLLOWED BY A
 CC UNIDIRECTIONAL GENE CONVERSION EVENT THAT REPLACES THE INFORMATION
 CC AT THE MAT LOCUS BY INFORMATION COPIED FROM EITHER OF THE TWO
 CC HOMOLOGOUS LOCI (HMR AND HMR') THAT RESIDE AT THE EXTREMITY OF THE
 CC CHROMOSOME III. ENDONUCLEASE EXPRESSION TAKES PLACE IN LATE G1
 CC JUST BEFORE CELLS ENTER S PHASE.
 CC MISCELLANEOUS: THE METAL-BINDING DOMAIN FORM ZINC-FINGERS THAT ARE
 CC INVOLVED IN BINDING OF THE DNA.
 CC -1- SIMILARITY: TO YEAST VMA1-DERIVED ENDONUCLEASE (VDE).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M4678; AAA34683.1; -

DR EMBL: X90957; CAA62447.1; -

DR EMBL: 274275; CAA98806.1; -

DR PIR: A25390; A25390.

DR HSP: P17255; 1YDE.

DR SGD: S0002386; HO.

KM Hydrolyase; Endonuclease; Zinc; Zinc-finger; DNA-binding.

FT SIMLAR 325 334 CONSERVED MOTIF FOR MITOCHONDRIAL MRNA

FT FT 189 189 A -> T (IN REF. 1).

FT FT 223 223 S -> G (IN REF. 1).

FT FT 405 405 S -> L (IN REF. 1).

FT FT 475 475 L -> H (IN REF. 1).

SQ SEQUENCE 586 AA; 66089 MW; 95771394D177823A CRC64;

Query Match 14.0%; Score 62; DB 1; Length 586;
Best Local Similarity 31.7%; Pred. No. 12;
Matches 20; Conservative 4; Mismatches 17; Indels 22; Gaps 5;

OY 30 KANSGRRRLC--CHRVPSNSTNLKHHVRLCKPCKLEPEPR-----LWVPG 76
DB 499 KMSG-KNVCAKRCY-----GRKKSQGH--CINCKYPERAEVKKAKDKGKLGITPE 549

OY 77 ALP 79
DB 550 GLP 552

RESULT 5
NTG3_HUMAN STANDARD: PRT; 632 AA.

AC P48066;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SODIUM- AND CHLORIDE-DEPENDENT GABA TRANSPORTER 3.
GN SLC6A11 OR GABT3 OR GAT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95179472; PubMed=7874447;
RA Borden L.A., Murali Dhar T.G., Smith K.E., Branchek T.A.,
RA Gluhowski C., Weisshank R.L.;
RT Cloning of the human homologue of the GABA transporter GAT-3 and
RT identification of a novel inhibitor with selectivity for this site.;
RL Recept. Channels 2:207-213(1994).
CC -1- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: WIDESPREAD DISTRIBUTION IN THE BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SODIUM-NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: S75989; AAB33570.1; -
DR EMBL: IPR000175; -
DR InterPro: IPR002982; -
DR Pfam: PF00209; SNF; 1.

DR PRINTS; PR00176; NANEUSMPORT.
DR PRINTS; PR01197; GAT3TRANSPORT.
DR PROSITE; PS00610; NA_NEUROTRAN_SYM_1; 1.
DR PROSITE; PS00754; NA_NEUROTRAN_SYM_2; 1.
DR PROSITE; PS02677; NA_NEUROTRAN_SYM_3; 1.
KM Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW Symport; Multigene family.
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79 1 (POTENTIAL).
FT TRANSMEM 87 106 2 (POTENTIAL).
FT TRANSMEM 131 151 3 (POTENTIAL).
FT DOMAIN 152 225 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 226 244 4 (POTENTIAL).
FT TRANSMEM 253 270 5 (POTENTIAL).
FT TRANSMEM 306 323 6 (POTENTIAL).
FT TRANSMEM 335 356 7 (POTENTIAL).
FT TRANSMEM 389 408 8 (POTENTIAL).
FT TRANSMEM 438 456 9 (POTENTIAL).
FT TRANSMEM 473 493 10 (POTENTIAL).
FT TRANSMEM 514 533 11 (POTENTIAL).
FT TRANSMEM 553 571 12 (POTENTIAL).
FT DOMAIN 572 632 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 632 AA; 70605 MW; D75C5B988B983278 CRC64;

Query Match 14.0%; Score 62; DB 1; Length 632;
Best Local Similarity 32.9%; Pred. No. 13;
Matches 24; Conservative 11; Mismatches 18; Indels 20; Gaps 5;

OY 3 LVISSLCIT-LLICFSIFTEGKRRPAKWSGRRRLCCHRVPSNSTNLKHHVRLCK 61
DB 536 LMAISLMICIPMICITVWKTEG-TLPER-----LQKITTP-STDLKMRG----- 598

OY 62 PCKLEPEPRIMV 74
DB 599 -KLGVSPPRMVTV 609

RESULT 6
MOG_HUMAN STANDARD: PRT; 247 AA.

AC Q16653; Q14855; Q13054; Q13055; Q92891; Q92892; Q92893; Q92894;
AC Q92895; Q93053; Q99605; Q00713; Q00714; Q00715;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN PRECURSOR.
GN MOG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC MEDLINE=95310943; PubMed=7790876;
RX HILTON A.A., SLAVIN A.J., HILTON D.J., BERNARD C.C.A.;
RT "Characterization of cDNA and genomic clones encoding human myelin
RT oligodendrocyte glycoprotein";
RL J. Neurochem. 65:305-318(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95054056; PubMed=7964757;
RA Pham-Dinh D., Allibaut B., Ruberg M., della Gaspera B.,
RA Nussbaum J.-L., Dautigny A.;
RT Characterization and expression of the cDNA coding for the human
RT myelin/oligodendrocyte glycoprotein.;
RL J. Neurochem. 63:2353-2356(1994).
RN [3]
RP SEQUENCE FROM N.A.


```
CC -1- SIMILARITY: BELONGS TO THE NOS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X97013; CA65736.1; -.
CC
CC DR HSSP: P29477; 2NOS.
CC
CC KM Oxidoreductase; NADP; FAD; FMN; Calmodulin-binding; Heme.
CC
CC FT NON_TER 1 1
CC
CC FT DOMAIN 139 159 CALMODULIN-BINDING (POTENTIAL).
CC NP_BIND 253 284 FMN (PYRIMIDINE PART) (BY SIMILARITY).
CC NP_BIND 398 409 FAD (ADP PART) (BY SIMILARITY).
CC NON_TER 470 470
CC
CC SQ SEQUENCE 470 AA; 53329 MW; 40B6717EE500B64D CRC64;

Query Match 13.7%; Score 60.5; DB 1; Length 470;
Best Local Similarity 22.5%; Pred. No. 15;
Matches 20; Conservative 6; Mismatches 32; Indels 31; Gaps 1;

OY 24 GRRPAKASGRRRLCCHRPSPNSTNKGHHVRLCKPKLE----- 66
DB 326 GAEKREAMDPLRHRYAVSCPDRTTALSAHSKAVLPMKLSKHNLQSSSSSTILV 385
OY 67 -----PEPLMVVPGALPOV 81
DB 386 ELERERSPEVMDFAFGDHYGFPPGNLPOL 414

RESULT 8
SM4B_MOUSE STANDARD; PRT; 782 AA.
ID 062179;
AC 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DI 30-MAY-2000 (Rel. 39, Last annotation update)
DE SEMAPHORIN 4B (SEMAPHORIN C) (SEMA C) (FRAGMENT).
GN SEM4B OR SEMAC OR SEMC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI; TISSUE-Brain;
RX MEDLINE=95267431; PubMed=7748561;
RA Puschel A.W., Adams R.H., Betz H.;
RT "Murne semaphorin D/collapsin is a member of a diverse gene family
RT and creates domains inhibitory for axonal extension.";
RL Neuron 14:941-948(1995).

-1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. LOW
CC LEVELS FOUND BETWEEN DAYS 10-12. EXPRESSION PEAKS ON DAY 13 WITH
CC MODERATE LEVELS FROM THEN UNTIL BIRTH.
CC
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC
CC -1- SIMILARITY: CONTRAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X85992; CA59984.1; -.
CC
```

```
DR MGD: I07559; Sem4b.
DR InterPro: IPR001627; -.
DR InterPro: IPR002165; -.
DR Pfam: PF01437; Plexin_repeat; 1.
DR Pfam: PF01403; Sem4; 1.
KW Transmembrane; Immunoglobulin domain; Multigene family; Neurogenesis;
KW development protein; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN 1 1 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 663 683 POTENTIAL.
FT DOMAIN 684 782 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 202 492 SEMA.
FT DOMAIN 548 608 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 703 726 PRO-RICH.
FT CARBOHYD 555 601 BY SIMILARITY.
FT CARBOHYD 12 12 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 782 AA; 86823 MW; 627A81FC8F87AC8 CRC64;

Query Match 13.7%; Score 60.5; DB 1; Length 782;
Best Local Similarity 27.3%; Pred. No. 24;
Matches 27; Conservative 10; Mismatches 23; Indels 39; Gaps 6;

OY 18 SIFSTEG-----KRRPAKASGRRRLCCHRPSPNSTNKGHHVRLCKPKLE----- 60
DB 473 SLPTGCDCLARDPYCAWTSACRLASLYPDLSRPTQDIEGASVKEICKNSSYKAR 532
OY 61 -----KPKCK--LEPE-----PEPLMVVPGPA 77
DB 533 FLVPGKPKQVQIQPNTVNTLACPLSLNATRLMWHNGA 571

RESULT 9
CD22_MOUSE STANDARD; PRT; 862 AA.
ID CD22_MOUSE
AC P35329;
DT 01-FEB-1994 (Rel. 28, Created)
DI 01-FEB-1994 (Rel. 28, Last sequence update)
DE B-CELL RECEPTOR CD22 PRECURSOR (LEU-14) (B-LYMPHOCYTE CELL ADHESION
DE MOLECULE) (BL-CAM).
GN CD22 OR LYB-8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D8A/2U; AND BALB/C; TISSUE-Liver;
RX MEDLINE=93315834; PubMed=8100843;
RA Law C.-L., Torres R.M., Sundberg H.A., Parkhouse R.M.,
RA Brannan C.I., Copeland N.G., Jenkins N.A., Clark E.A.;
RT "Organization of the murine Cd22 locus. Mapping to chromosome 7 and
RT characterization of the two alleles.";
RL J. Immunol. 151:175-187(1993).

-1- FUNCTION: MEDIATES B-CELL B-CELL INTERACTIONS. MAY BE INVOLVED IN
CC THE LOCALIZATION OF B-CELLS IN LYMPHOID TISSUES. BINDS STYLYLATED
CC GLYCOPROTEINS; ONE OF WHICH IS CD45.
CC
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, CD22-ALPHA AND CD22-BETA (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC
CC -1- TISSUE SPECIFICITY: B-LYMPHOCYTES.
CC
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 4 C2-LIKE AND ONE V-LIKE DOMAINS.
CC
```


CC -1- SIMILARITY: TO MYELIN-ASSOCIATED GLYCOPROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, L16928; AAA02562.1; -
 CC MGD; MGI:88322; Cd22.
 CC InterPro: IPR003006; -
 CC Pfam: PF00047; 19; 6.
 CC Glycoprotein; Cell adhesion; Transmembrane; Signal; B-cell;
 CC Immunoglobulin domain; Alternative splicing.
 CC SIGNAL 1 21
 CC CHAIN 1 21
 CC DOMAIN 22 862
 CC TRANSMEM 703 721
 CC DOMAIN 722 862
 CC DOMAIN 22 142
 CC DOMAIN 158 235
 CC DOMAIN 265 331
 CC DOMAIN 361 418
 CC DOMAIN 450 506
 CC DOMAIN 537 593
 CC DOMAIN 624 681
 CC DISULFID 41 171
 CC DISULFID 46 106
 CC DISULFID 165 229
 CC DISULFID 272 324
 CC DISULFID 368 411
 CC DISULFID 457 499
 CC DISULFID 544 586
 CC DISULFID 631 674
 CC CARBOHYD 105 105
 CC CARBOHYD 116 116
 CC CARBOHYD 139 139
 CC CARBOHYD 168 168
 CC CARBOHYD 255 265
 CC CARBOHYD 275 275
 CC CARBOHYD 378 378
 CC CARBOHYD 408 408
 CC CARBOHYD 460 460
 CC CARBOHYD 561 561
 CC CARBOHYD 589 589
 CC VARIANT 15 15
 CC VARIANT 19 19
 CC VARIANT 76 76
 CC VARIANT 82 86
 CC VARIANT 90 91
 CC VARIANT 94 94
 CC VARIANT 102 102
 CC VARIANT 173 173
 CC VARIANT 179 179
 CC VARIANT 186 186
 CC VARIANT 190 190
 CC VARIANT 192 192
 CC VARIANT 236 236
 CC VARIANT 241 241
 CC VARIANT 244 244
 CC VARIANT 616 616
 CC VARIANT 714 714
 CC VARIANT 787 787
 CC VARIANT 808 808
 CC SEQUENCE 862 AA; 96582 MW; A7662D6E87038E83 CRC64;

Query Match 13.7%; Score 60.5; DB 1; Length 862;
 Best Local Similarity 29.6%; Pred. No. 26;
 Matches 16; Conservative 11; Mismatches 20; Indels 7; Gaps 2;

OY 28 PAKMSGRRTICCHRVSPNSNLIKHHVRLCKPCKLEPERLWVPGALPOV 81
 ID 355 PSFAEEGQVELCESLASPSATNTWYHNR--KPIPGTQELK-----RIBKV 401
 Db
 RESULT 10
 ID URT2_DESRO STANDARD; PRT; 477 AA.
 AC P15638;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SALIVARY PLASMINOGEN ACTIVATOR ALPHA 2 PRECURSOR (EC 3.4.21.68) (DSFA
 DE ALPHA-2) (BAT-PA) (T-PLASMINOGEN ACTIVATOR).
 OS Desmodus rotundus (Vampire bat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
 OC Desmodontinae; Desmodus.
 OX NCBI_TaxID=9430;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Salivary gland;
 RX MEDLINE-92039036; PubMed-1937019;
 RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,
 RA Alagon A., Donner P., Schleuning W.D.;
 RT "The plasminogen activator family from the salivary gland of the
 RT vampire bat Desmodus rotundus: cloning and expression.";
 RL Gene 105:229-237(1991).
 RN [2]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC TISSUE-Salivary gland;
 RX MEDLINE-90036867; PubMed-2509450;
 RA Gardell S.J., Duong L.T., Diehl R.E., York J.D., Hare T.R.,
 RA Register R.B., Jacobs J.W., Dixon R.A.F., Friedman P.A.;
 RT "Isolation, characterization, and cDNA cloning of a vampire bat
 RT salivary plasminogen activator.";
 RL J. Biol. Chem. 264:17947-17952(1989).
 RN [3]
 RP CHARACTERIZATION.
 RC MEDLINE-93393059; PubMed-1309059;
 RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,
 RA Kraetzschmar J., Haendler B., Langer G., Baldus B., Wilt W.,
 RA Donner P.;
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common
 RT vampire bat): unique fibrin specificity.";
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992)
 CC -1- FUNCTION: PROBABLY ESSENTIAL TO SUPPORT THE FEEDING HABITS OF THIS
 CC EXCLUSIVELY HAEMATOPHAGOUS ANIMAL. PROBABLE POTENT THROMBOLYTIC
 CC AGENT.
 CC -1- CATALYTIC ACTIVITY: SPECIFIC CLEAVAGE OF ARG-1-VAL BOND IN
 CC PLASMINOGEN TO FORM PLASMIN.
 CC -1- ENZYME REGULATION: ACTIVITY TOWARD PLASMINOGEN IS STIMULATED IN
 CC THE PRESENCE OF FIBRIN I.
 CC -1- SUBUNIT: MONOMER.
 CC -1- DOMAIN: THE FIBRONECTIN TYPE-1 DOMAIN MEDIATES BINDING TO FIBRIN,
 CC AND THE KRINGLE DOMAIN APPARENTLY MEDIATES FIBRIN-INDUCED
 CC STIMULATION OF ACTIVITY.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-1 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE REGION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M63988; AAA31593.1; -
 CC EMBL, J05082; AAA31596.1; -

DR PIR: A34369; A34369.
 DR HSP: P00750; 1RPF.
 DR MEROPS: S01.232; -.
 DR InterPro: IPR0000001; -.
 DR InterPro: IPR000083; -.
 DR InterPro: IPR000561; -.
 DR InterPro: IPR001254; -.
 DR InterPro: IPR001314; -.
 DR Pfam: PF00008; EGF_1.
 DR Pfam: PF00039; fnl; 1.
 DR Pfam: PF00051; kringle; 1.
 DR Pfam: PF00088; trypsin; 1.
 DR PRINTS: PRO0018; KRYOTRYP.
 DR PRINTS: PRO022; CHYMOTRYPSIN.
 DR PROSITE: PS00021; KRINGLE_1; 1.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01253; FIBRONECTIN_1; 1.
 DR PROSITE: PS0070; KRINGLE_2; 1.
 DR Plasminogen activation; Hydrolyase; Serine protease; Glycoprotein;
 KW kringle; EGF-like domain; Signal; Multigene family.
 FT SIGNAL 1 36
 FT CHAIN 37 477
 FT DOMAIN 40 82
 FT DOMAIN 83 121
 FT DOMAIN 128 209
 FT DOMAIN 225 477
 FT ACT_SITE 272 272
 FT ACT_SITE 321 321
 FT ACT_SITE 428 428
 FT DISULFID 42 72
 FT DISULFID 70 79
 FT DISULFID 87 98
 FT DISULFID 92 109
 FT DISULFID 111 120
 FT DISULFID 128 209
 FT DISULFID 149 191
 FT DISULFID 180 204
 FT DISULFID 214 345
 FT DISULFID 257 273
 FT DISULFID 265 334
 FT DISULFID 359 434
 FT DISULFID 391 407
 FT DISULFID 424 452
 FT CARBOHYD 185 185
 FT CARBOHYD 398 398
 FT CONFLICT 403 403
 FT CONFLICT 417 417
 FT CONFLICT 435 435
 SO SEQUENCE 477 AA; 53719 MW; 174865550E5077C CMC64;

Query Match 13.6%; Score 60; DB 1; Length 477;
 Best Local Similarity 22.4%; Pred.No. 17;
 Matches 19; Conservative 11; Mismatches 17; Indels 38; Gaps 4;

OY 7 SLLCILLCLPSIFS-----TEGKR-----RPAKAW-----SGRR 36
 DB 7 TKLLCVLLGAVFSLPROETRYROLARSGVACRDEKTOITVQOQESWLRPEVRSKR 66
 OY 37 TRLC-----CHRVSPNSTNLK 53
 DB 67 VEHCRDRLGACHTVYKSCSELR 91

RESULT 11
 TSP2_MOUSE STANDARD; PRT: 1172 AA.
 AC 003350;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THBS2 OR TSP2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 CC NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RT Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development."
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [2]
 RP SEQUENCE OF 1-873 FROM N.A.
 RX MEDLINE=91302287; PubMed=1712771;
 RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
 RA Dixit V.M.;
 RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
 genome."
 RL J. Biol. Chem. 266:12821-12824(1991).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 WVEC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L07803; AAA53064.1; -.
 DR EMBL: M64866; AAA40432.1; -.
 DR PIR: A42587; A42587.
 DR PIR: A39851; A39851.
 DR HSP: P00740; 1TXA.
 DR MGD: MGI:98738; Thbs2.
 DR InterPro: IPR000561; -.
 DR InterPro: IPR000884; -.
 DR InterPro: IPR001007; -.
 DR Pfam: PF00008; EGF_2; 1.
 DR Pfam: PF00090; tsp_1; 3.
 DR Pfam: PF00093; wvc; 1.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS50092; TSP1; 3.
 DR PROSITE: PS01208; WVEC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1172
 FT DOMAIN 19 232
 FT DOMAIN 318 375
 FT DOMAIN 381 432
 FT DOMAIN 437 493
 FT DOMAIN 494 548
 FT DOMAIN 549 589
 FT DOMAIN 590 647
 FT DOMAIN 648 692
 FT DOMAIN 725 760
 FT DOMAIN 761 783
 FT DOMAIN 784 819
 FT DOMAIN 820 842

POTENTIAL.
 THROMBOSPONDIN 2.
 HEPARIN-BINDING (POTENTIAL).
 WVEC.
 TSP TYPE-1 1.
 TSP TYPE-1 2.
 TSP TYPE-1 3.
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 CALCIUM-BINDING (POTENTIAL).
 TSP TYPE-3 1.
 TSP TYPE-3 2.
 TSP TYPE-3 3.
 TSP TYPE-3 4.

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X00588; CAA25240.1; -
 DR EMBL: X06370; CAA29668.1; -
 DR EMBL: X06653; CAA25282.1; -
 DR EMBL: M38425; AAA63171.1; -
 DR EMBL: M11234; AAA52370.1; -
 DR PIR: A00641; G0HHE.
 DR PIR: A00642; G0HHE2.
 DR PIR: A23062; A23062.
 DR HSSP: P11362; 1F61.
 DR SWISS-2DPAGE: P00533; HUMAN.
 DR MIM: 131550; -
 DR InterPro: IPR000494; -
 DR InterPro: IPR000719; -
 DR InterPro: IPR001245; -
 DR InterPro: IPR002174; -
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep.L.domain; 2.
 DR Pfam: PF00069; pkinase; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM_1.
 DR Transmembrane: Glycoprotein; Duplication; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 645
 FT TRANSMEM 646 668
 FT DOMAIN 669 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1025 1071
 FT DOMAIN 712 979
 FT NP_BIND 718 726
 FT BINDING 745 745
 FT ACET_SITE 837 837
 FT MOD_RES 678 678
 FT MOD_RES 1092 1092
 FT MOD_RES 1110 1110
 FT MOD_RES 1172 1172
 FT MOD_RES 1197 1197
 FT CARBOHYD 128 128
 FT CARBOHYD 175 175
 FT CARBOHYD 196 196
 FT CARBOHYD 352 352
 FT CARBOHYD 361 361
 FT CARBOHYD 413 413
 FT CARBOHYD 444 444
 FT CARBOHYD 528 528
 FT CARBOHYD 568 568
 FT CARBOHYD 603 603
 FT CARBOHYD 623 623
 FT CONFLICT 540 540
 SO SEQUENCE 1210 AA; 134277 MW; DBA2A50B4EFBBD2 CRC64;

Query Match 13.5%; Score 59.5; DB 1; Length 1210;
 Best Local Similarity 26.7%; Pred. No. 45;
 Matches 27; Conservative 8; Mismatches 27; Indels 39; Gaps 6;
 Oy 15 LCFS-----IFSTGRRRPAKMSGRRLC-----GRVSPSTNKGHHVRLCK 61
 Db 469 LCYANTIMWKLFSTGSK--TKIISNGENSCRAKQVCHALCSPE--GCGWGPEDCV 524
 Oy 62 PCK-----LEPRRLMV-----PGALPQ 80
 Db 525 SCRNVSRGRCVDCNLCLEGPREFVENSECTIOCHPECLPQ 565

RESULT 13
 ID ATP8_LATCH STANDARD; PRT; 55 AA.
 AC 003168;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (ATPase SUBUNIT 8) (A6L).
 OS Lattimeria OR ATP8.
 GN Lattimeria chalumnae (Lattimeria) (Coelacanth).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Coelacanthiformes; Coelacanthidae; Lattimeria.
 OX NCBI_TaxID=7897;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zardoya R., Meyer A.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
 CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPase COMPLEX.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC -1- SIMILARITY: BELONGS TO THE ATPase PROTEIN 8 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U82228; AAC60322.1; -
 DR InterPro: IPR001421; -
 DR Pfam: PF00895; ATP-synt-8; 1.
 KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
 FT TRANSMEM 4 24
 FT SEQUENCE 55 AA; 6569 MW; 7EE36319E8AF825B CRC64;

Query Match 13.3%; Score 59; DB 1; Length 55;
 Best Local Similarity 34.3%; Pred. No. 3.3;
 Matches 23; Conservative 7; Mismatches 15; Indels 22; Gaps 5;

Oy 7 SSLCILLCISISTEKKRRPAKMSGRRLCCHRVPS--PNSTNKGHHVRLCKPKL 65
 Db 7 SPWLLILFSLVLTFL--MLPSK-----TOL--HTPPNPSTQNM-----CKQ 45
 Oy 66 EPEPRLW 72
 Db 46 EPEPRLW 52

RESULT 14
 ID FXH1_HUMAN STANDARD; PRT; 365 AA.
 AC 075593;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE FORKHEAD BOX PROTEIN H1 (FORKHEAD ACTIVITY SIGNAL TRANSDUCER 1)
 DE (FAST-1).
 GN FOXH1 OR FAST1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98367554; PubMed=9702198;
 RA Zhou S., Zavel L., Lengauer C., Kinzler K.W., Vogelstein B.;
 RT "Characterization of human FAST-1, a TGF beta and activin signal
 transducer.";

```

RL Mol. Cell 2:121-127(1998).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES AN ACTIVIN RESPONSE
CC ELEMENT (ARE). RECOGNIZES AND BINDS TO THE DNA SEQUENCE 5'-
CC TGTGTGTTGGTAT-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF076292; AAC34303.1; .
DR HSSP; 063245; 2HFH.
DR MIM; 603621; .
DR InterPro: IPR001766; .
DR Pfam; PF00250; FORK_head; 1.
DR PRINTS; PS00053; FORKHEAD.
DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
DR Transcription regulation; Activator; DNA-binding; Nuclear protein.
KW DNA_BIND
FT 32 128 FORK-HEAD.
SO SEQUENCE 365 AA; 39257 MW; 27A13F39C089F722 CRC64;

Query Match 13.3%; Score 59; DB 1; Length 365;
Best Local Similarity 35.4%; Pred. No. 18;
Matches 17; Conservative 3; Mismatches 14; Indels 14; Gaps 3;

QY 39 LCCHRVSPSPSTNLKGHVR--LCRPCKLEPPRLM-----VYPG 76
DB 219 LC-----PLPGPTVEGETYQGAIGPSTLSPPEPRAPLHLGLGTAVPG 262

RESULT 15
ID GAT5_CHICK STANDARD; PRT; 391 AA.
AC P43692;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRANSCRIPTION FACTOR GATA-5 (GATA BINDING FACTOR-5).
GN GATA5.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RX MEDLINE=94365018; PubMed=8083222;
RA Lavertiere A.C., Macneil C., Mueller C., Poelmann R.E.,
RA Burch J.B.E., Evans T.;
RA "GATA-4/5/6, a subfamily of three transcription factors transcribed
RT in developing heart and gut.";
RL J. Biol. Chem. 269:23177-23184(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH AND SMALL INTESTINE,
CC LOWER LEVELS IN HEART, LUNG AND SPLEEN. VERY LOW LEVELS IN LIVER
CC AND OVARY.
CC -1- DEVELOPMENTAL STAGE: INITIALLY TRANSCRIBED IN THE CARDIAC CRESCENT
CC PRIOR TO FORMATION OF THE PRIMORDIAL HEART TUBE. FOLLOWING
CC FORMATION OF THE PRIMITIVE HEART, PRESENT IN BOTH ENDOCARDIUM AND
CC MYOCARDIUM AS WELL AS IN OTHER LATERAL PLATE DERIVATIVES. ALSO
CC TRANSCRIBED IN THE PRIMITIVE EMBRYONIC GUT AND IN LATE STAGE
CC EMBRYOS IS SPONTANEOUSLY UP-REGULATED IN DISTINCT SEGMENTS OF
CC GASTROINTESTINAL EPITHELIA AS THEY UNDERGO TERMINAL

```

```

CC DIFFERENTIATION.
CC -1- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U11888; AAA57504.1; .
DR HSSP; P17678; 1GAN.
DR InterPro: IPR000679; .
DR Pfam; PF00320; GATA; 2.
DR PRINTS; PR00619; GATA2NFINGER.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 2.
DR Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT 186 210 GATA-TYPE.
FT ZN_FING 239 GATA-TYPE.
SO SEQUENCE 391 AA; 41858 MW; EF8A283111824260 CRC64;

Query Match 13.3%; Score 59; DB 1; Length 391;
Best Local Similarity 32.8%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 28; Indels 14; Gaps 4;

QY 16 CFSIFTEGRKRRPAKAW-----SGRRRLCCHRPSPNST-----NLKGHVRCLKPC----- 63
DB 210 CGLYHKMNSINRLPKQKRSLSSRRAGLCCTNCHTNTTLTMRNRAGEPV--CNAGGLYM 267
QY 64 KLEPEPR 70
DB 268 KLHGVR 274

```

Search completed: June 3, 2001, 04:38:06
Job time: 206 sec

A:Reference number: A92935; MUID:87169717
 A:Accession: D27216
 A:Molecule type: DNA
 A:Residues: 1-462 <MES>
 A:Cross-references: EMBL:X04650; NID:959801; PIDN:CAA2339.1; PID:959808
 A:Experimental source: strain AD169
 R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Holsnell, T.;
 M.; Bairelli, B.G.
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
 A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus
 A:Reference number: S09749; MUID:90269039
 A:Accession: S09943
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-462 <CHE>
 A:Cross-references: EMBL:X17403; NID:959591; PIDN:CAA35261.1; PID:g1780960
 A:Experimental source: strain AD169
 A:Note: this sequence was submitted to the EMBL Data Library, December 1989
 C:Genetics:
 A:Gene: HHRP4
 C:Superfamily: cytomegalovirus HHRP4 protein

Query Match 15.0%; Score 66.5; DB 1; Length 462;
 Best Local Similarity 25.8%; Pred. No. 8.6;
 Matches 23; Conservative 7; Mismatches 42; Indels 17; Gaps 3;

OY 4 LVLSLILLCILLC--FSIFTEGRRPAKMS-----GRTRLCCH--VPS 46
 DB 258 LCVDLVLCVLLALLLELVPMENVRHPLTFRRVALSPSTSKVDRAVKLCRRHFGLP 317

OY 47 PNSTNLKHHVRLCPCKLEPPRIWVP 75
 DB 318 PPSVAPGEEKELPQAALSPPLTWSP 346

RESULT 3
 H71417
 Cytochrome P450 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: Columbia
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
 C:Accession: H71417
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; Giel
 vanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puldomenich
 erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Falme, K.; Benes, V.; Rechman, S.; Ans
 C.; Chaiwatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 A:Reference number: A71400; MUID:98121113
 A:Accession: H71417
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-515 <BEV>
 A:Cross-references: GB:297338; NID:92244870
 C:Genetics:
 A:Map position: 4COP9-463845
 C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
 F:308-476/Domains: cytochrome P450 homology <P45>

Query Match 14.8%; Score 65.5; DB 2; Length 515;
 Best Local Similarity 30.6%; Pred. No. 12;
 Matches 19; Conservative 9; Mismatches 11; Indels 23; Gaps 3;

OY 1 MRLVLSLILLCIL-LGCSIF-----STEGRRPAKMSRRRLCCHRPSPNSTNLK 53
 DB 1 MAVLIIFILCLLSFLCYSLEFMKPKDSRDGRDLP-----PSPSPPLPII 44

OY 54 GH 55
 11

DB 45 GH 46

RESULT 4
 A36302
 Submaxillary protein SMR1 precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 02-Jun-2000
 C:Accession: A36302; A31347; S20792
 R:Rosinski-Chupin, I.; Rougeon, F.
 DNA Cell Biol. 9, 553-559, 1990
 A:Title: The gene encoding SMR1, a precursor-like polypeptide of the male rat submaxi
 A:Reference number: A36302; MUID:91103875
 A:Accession: A36302
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <ROS1>
 A:Cross-references: EMBL:J04109; EMBL:X52467; NID:957256; PIDN:CAA36705.1; PID:g57257
 R:Rosinski-Chupin, I.; Tronk, D.; Rougeon, F.
 Proc. Natl. Acad. Sci. U.S.A. 85, 8553-8557, 1988
 A:Title: High level of accumulation of a mRNA coding for a precursor-like protein in
 A:Reference number: A31347; MUID:89042220
 A:Accession: A31347
 A:Molecule type: mRNA
 A:Residues: 1-113; 'H', 115-146 <ROS2>
 A:Cross-references: GB:M63112; GB:J04109; NID:9206997; PIDN:AAA42153.1; PID:g206998
 C:Genetics:
 A:Introns: 18/3
 C:Superfamily: proline-rich peptide P-B
 C:Keywords: glycoprotein
 F:1-18/Domains: signal sequence #status predicted <SIG>
 F:19-146/Product: submaxillary protein SMR1 #status predicted <MAT>
 F:129,136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.1%; Score 62.5; DB 2; Length 146;
 Best Local Similarity 34.9%; Pred. No. 9;
 Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6;

OY 1 MRLVLSLILLCILLCFSIFTEGRRPAKMSGRTR--LCCHRY--PSPNSTNLK 54
 DB 1 MKSLYIFGLWILNACFO--SGEYGRPRRQHNPRRODPSTLPHYLGLODPNGOI-- 56

OY 55 HHVRLCKPCKLEPPRIWVP 76
 DB 57 GVTITPLNQP-PRVLVNLPG 77

RESULT 5
 S59301
 homothallic switching endonuclease - Yeast (Saccharomyces cerevisiae)
 N:Alternate names: HO endonuclease; homothallic protein; protein D0827; protein YDL2
 C:Species: Saccharomyces cerevisiae
 C:Date: 08-Jul-1995 #sequence_revision 01-Dec-1995 #text_change 29-Oct-1999
 C:Accession: S59301; A25390; S67790
 R:Raven, D.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: S59301
 A:Accession: S59301
 A:Molecule type: DNA
 A:Residues: 1-586 <RAV>
 A:Cross-references: EMBL:X90957; NID:9984693; PIDN:CAA62447.1; PID:9984694
 R:Russell, D.W.; Jensen, R.; Zollier, M.J.; Burke, J.; Errede, B.; Smith, M.; Herskowi
 Mol. Cell. Biol. 6, 4281-4294, 1986
 A:Title: Structure of the Saccharomyces cerevisiae HO gene and analysis of its upstre
 A:Reference number: A25390; MUID:87089786
 A:Accession: A25390
 A:Molecule type: DNA
 A:Residues: 1-188; 'T', 190-222; 'G', 224-404; 'T', 406-474; 'H', 476-586 <RUS>
 A:Cross-references: EMBL:M14678; NID:9171697; PIDN:AAA34683.1; PID:g171698
 R:Rasmussen, S.W.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67778

A:Accession: S67790
 A:Molecule type: DNA
 A:Residues: 1-586 <RBS>
 A:Cross-references: EMBL:Z74275; NID:g1431382; PIDN:CAA98806.1; PID:e253273; PID:g143138
 A:Experimental source: strain S286C
 C:Genetics:
 A:Gene: SGD:HO
 A:Cross-references: SGD:S0002386; MIPS:YDL227c
 A:Map position: 4L
 C:Function:
 A:Description: site-specific endonuclease that cleaves a site in the MAT locus on chromo
 C:Keywords: DNA binding; nucleus; zinc finger

Query Match 14.0%; Score 62; DB 2; Length 586;
 Best Local Similarity 31.7%; Pred. No. 32;
 Matches 20; Conservative 4; Mismatches 17; Indels 22; Gaps 5;

QY 30 KAMSGRRRLC--CHRPSPNSTNLKGHWRLCKKCKLEPPR-----LWVPG 76
 DB 499 KDMSG-KNRVCARCY-----GRYKFSQHH--CLNCKYVPEARVKKAKDKGKLGITPE 549

QY 77 ALP 79
 DB 550 GLP 552

RESULT 6
 153030
 submaxillary protein SMRI precursor - black rat
 C:Species: Rattus rattus (black rat, roof rat)
 A:Note: This sequence is derived from an apparently genuine specimen of this problematic
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Jun-2000
 C:Accession: 153030
 R: Singer, M.; Courty, Y.; Rougeon, F.
 DNA Cell Biol. 14, 137-144, 1995
 A:Title: Recent evolution of genes encoding the prohormone-like protein SMRI in the rat
 A:Reference number: 153030; MID:95169272
 A:Accession: 153030
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-146 <RES>
 A:Cross-references: EMBL:X77819; NID:g732921; PIDN:CAA54834.1; PID:g732922
 C:Genetics:
 A:Gene: VCS-alpha1
 C:Superfamily: proline-rich peptide P-B
 C:Keywords: glycoprotein
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-146/Product: submaxillary protein SMRI #status predicted <MAT>
 F:129,136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.9%; Score 61.5; DB 2; Length 146;
 Best Local Similarity 33.7%; Pred. No. 12;
 Matches 28; Conservative 10; Mismatches 32; Indels 13; Gaps 5;

QY 1 MRLIVSLICILLCFSTFTEGKRRPAKAMSGRR-----TRLCCHRPSPNSTNLKG 54
 DB 1 MKRPLYLFGILWILGCFQ--SGEGDRPRRQHNLRROTSTLHYLGLQDPDVMQMR 57
 QY 55 HHVRLCKPCKLEPPRLMW-VPG 76
 DB 58 --VITIPNLDP-PRVLVNLPG 77

RESULT 7
 E70253
 hypothetical protein BBR19 - Lyme disease spirochete plasmid K/1p36
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: E70253
 R: Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kierlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MID:98065943
 A:Accession: E70253
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-211 <REF>
 A:Cross-references: GB:AE000788; NID:g2690123; PIDN:AAAC6172.1; PID:g2690164; TIGR:BB
 A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid

Query Match 13.9%; Score 61.5; DB 2; Length 211;
 Best Local Similarity 26.9%; Pred. No. 16;
 Matches 18; Conservative 13; Mismatches 21; Indels 15; Gaps 2;

QY 1 MRLIVSLICILLCFSTFTEGKRR-----PAKAMSGRRRLCCHRPSPNSTNLK 53
 DB 1 MKXYIIMSLICILLCFSTFTEGKRR-----NPINKENIDTEK 52

QY 54 GHVRLC 60
 DB 53 GTNTTLC 59

RESULT 8
 S58075
 probable olfactory receptor tpcr18 - rat (fragment)
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 26-Aug-1999
 C:Accession: S58075
 R: Vandenheghen, P.; Schumann, S.; Vassart, G.; Parmentier, M.
 submitted to the EMBL Data Library, July 1995
 A:Description: Male germ cells from several mammalian species express a specific repe
 A:Reference number: S57995
 A:Accession: S58075
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-154 <VAN>
 A:Cross-references: EMBL:X89702; NID:g902720; PIDN:CAA61849.1; PID:g902721
 C:Superfamily: olfactory receptor OR14

Query Match 13.8%; Score 61; DB 2; Length 154;
 Best Local Similarity 34.9%; Pred. No. 14;
 Matches 15; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 3 LTVLSLILLCFSTFTEGKRRPAKAMSGRRRLCCHRPV 45
 DB 87 LLITFSYSCIVCSILQIRSAEGRRRRASTCSAHLTALLFLYMP 129

RESULT 9
 J00137
 hypothetical 30.1K protein - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 07-Jun-1996
 R: Kato, J.; Chu, L.; Kitano, K.; Devault, J.D.; Kimbara, K.; Chakrabarty, A.M.; Misra
 Gene 84, 31-38, 1989
 A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in
 A:Reference number: J00137; MID:90108714
 A:Accession: J00137
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-261 <KAT>
 A:Note: 3-Met could also be the initiator
 C:Genetics:
 A:Start codon: GTC

OM of: US-09-599-087-5 to: N_Geneseq_0401.* out_format : pfs
 Date: Jun 3, 2001 5:15 AM

About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Comugen Ltd.

Command line parameters:

-MODEL-frame+P2n.model -DEV-rlp
 -O-cgn2_1/USPTO.spool/US09599087/runat_02062001_140345_25633/app_query.fasta.1.138
 -DB-N_Geneseq_0401 -QFM-fastap -SUFFIX-rng -GAPOP-12.000
 -GAPEXT-4.000 -MINMATCH-0.100 -LOOPEL-0.000 -LOOPEXT-0.000
 -GAPOP-4.500 -OGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500
 -FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500
 -DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-Biosum62
 -TRANS-human40.cdi -LIST-45 -DOCALLIGN-200 -THR_SCORE-pct
 -THR_MAX-100 -THR_MIN-0 -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs
 -NOR-ext -MINLEN-0 -MAXLEN-200000000
 -USER-US09599087 -CGN1_1.175 -NCPU-6 -ICPU-3 -LONGLOG -NO_XLPRY
 -WAIT -THREADS-1

Search information block:

Query: US-09-599-087-5
 Query length: 81
 Database: N_Geneseq_0401.*
 Database sequences: 678276
 Database length: 291890651
 Search time (sec): 120.320000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A72230				400.00	752.73 5.5e-34
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A72224				400.00	752.68 5.5e-34
/SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:057015				86.50	141.58 6.02
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C98948				85.00	141.20 6.32
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:T27617				85.00	139.61 7.75
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:T27616				85.00	139.35 8.00
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:T18996				85.00	139.35 8.00
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:T18996				85.00	139.35 8.00
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:X76729				84.50	134.63 14.67
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:F18164				81.00	140.29 7.09
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C74372				79.00	127.97 34.43
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:059506				79.00	143.30 4.83
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:T77840				79.00	129.42 28.62
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:T77838				79.00	127.08 38.62
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:X23057				79.00	117.82 126.70
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:X21355				79.00	117.82 126.70
/SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:C90314				79.00	117.82 126.70
/SID2/gcgdata/geneseq/geneseqn/NA1995.DAT:088760				78.00	127.06 38.72
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:T77634				76.50	126.60 41.06
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C77789				76.50	124.53 53.55
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C93400				75.50	121.69 77.12
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:X53300				74.50	87.39 6.3e+03
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:X22704				74.00	115.14 178.59
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A93118				73.50	119.68 99.76
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A93118				73.00	120.10 94.53
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A67275				72.50	128.57 31.90
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:T12815				72.00	131.55 21.75
/SID2/gcgdata/geneseq/geneseqn/NA1998.DAT:V27207				72.00	117.50 132.01
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C86435				72.00	113.42 222.62
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C86435				72.00	113.42 222.62
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C86411				72.00	111.06 301.41
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:C51033				72.00	107.43 480.15
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:T60073				72.00	106.72 526.13
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A67288				71.50	130.73 24.17
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C53213				71.50	127.28 37.65
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A51306				71.50	121.73 76.67
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C53204				71.50	118.38 117.90
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:289988				71.50	115.03 181.14
/SID2/gcgdata/geneseq/geneseqn/NA1998.DAT:V59986				71.50	111.72 276.76

/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A51295 + 71.50 86.10 7.4e+03 31328
 /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A81458 - 71.50 81.47 1.3e+04 49767
 /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:F21613 + 71.50 69.05 6.4e+04 172325
 /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:F21612 + 71.50 61.97 1.5e+05 349980
 /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:C99231 + 71.00 126.82 39.92 485

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:A72230

seq_documentation_block:

ID A72230 standard; cDNA; 797 BP.

AC A72230;

DT 06-DEC-2000 (first entry)

DE Human CASB gene partial cDNA, SEQ ID NO:7.

KM Human; CASB gene; overexpression; colon tumour-associated antigen;

KM expressed sequence tag; EST; colon cancer; tumour; autoimmune disease;

KM diagnosis; disease susceptibility; prophylaxis; genetic vaccine;

KM gene therapy; ss.

OS Homo sapiens.

PN WO20043509-A2.

PD 27-JUL-2000.

PF 17-JAN-2000; 2000MO-EP00346.

PR 19-JAN-1999; 99GB-0001078.

PR 29-JAN-1999; 99GB-0002090.

PR 01-FEB-1999; 99GB-0002163.

PR 01-FEB-1999; 99GB-0002168.

PR 01-FEB-1999; 99GB-0002169.

PR 07-APR-1999; 99GB-0007901.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Vinals-Bassols C;

WPI, 2000-482912/42.

New isolated polynucleotide useful for diagnosis and/or treatment of

colon cancer and autoimmune disease -

PS Disclosure: Page 35-36; 41pp; English.

Sequences A72230-A72235 represent human CASB gene partial cDNA

sequences which are derived from expressed sequence tags (ESTs). The

invention relates to human CASB cDNA sequences CASB611, CASB500, CASB501,

CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and also to

these human CASB partial cDNA sequences. Expression of the human

CASB genes (with the exception of CASB611) is associated with colon

tumours, and the encoded proteins (sequences not given in the

specification) represent colon tumour-associated antigens. The cDNA

sequences may be used in diagnosing the presence or activity of CASB genes.

Such diseases include autoimmune diseases and especially colon cancer.

The nucleic acid sequences may also be used in genetic vaccines for the

prophylaxis or therapeutic treatment of colon cancer and autoimmune

diseases.

Sequence 797 BP; 164 A; 180 C; 254 G; 199 T; 0 other;

alignment_scores:

Quality: 400.00 Length: 77
 Ratio: 5.263 Gaps: 1
 Percent Similarity: 98.701 Percent Identity: 97.403

alignment_block:

US-09-599-087-5 x A72230/rev

Align seg 1/1 to reverse of: A72220 from: 1 to: 797

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
765 ATGAGGCTTACGCTTTCAGGCTGCTGTATCTGCTTCTGCTT 716
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.AlaTppSer 33
|||||
715 CTCATCTTCTCCACAGAGGGAAGGCGCTCCTGCCACAGCGCTGCTCA 666
34 G1ArgArgThrArgLeuCysCysHisArgValProSerProAsnSerTh 50
|||||
665 GGCAGAGAGAACCAAGGCTCTGCTGCCACCGAGTCCCTAGCCCACTCAAC 616
50 rAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGlu 67
|||||
615 AAACCTGAAGAGACATCATGTGAGGCTGTAAACCATGCAACCTTGAGC 566
67 roGluProArgLeuTTPValValProGly 76
|||||
565 CAGAGCCCGCCCTTGGTGGTGGCTGGG 537

seq_name: /SID2/gcdata/geneseq/geneseqn/NA2000.DAT:A72224

seq_documentation_block:
ID A72224 standard; cDNA; 801 BP.
AC A72224;
XX
XX 06-DEC-2000 (first entry)
DE Human CASB611 cDNA.
XX
XX Human: CASB611: colon-specific expression; expressed sequence tag;
KW EST: colon cancer; tumour; autoimmune disease; diagnosis;
KW disease susceptibility; prophylaxis; genetic vaccine; gene therapy; ss.
XX
XX Homo sapiens.
PN WO20043509-A2.
XX
XX 27-JUL-2000.
PD
XX
XX 17-JAN-2000; 2000WO-EP00346.
XX
XX 19-JAN-1999; 99GB-0001078.
PR 29-JAN-1999; 99GB-0002090.
PR 01-FEB-1999; 99GB-0002163.
PR 01-FEB-1999; 99GB-0002168.
PR 01-FEB-1999; 99GB-0002169.
PR 07-APR-1999; 99GB-0007901.
XX
XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
PA
XX
XX Vinals-Bassols C;
PI
XX
XX WPI; 2000-482912/42.
DR
XX
XX New isolated polynucleotide useful for diagnosis and/or treatment of
PT colon cancer and autoimmune disease -
XX
XX
XX Claim 3; Page 34; 41pp; English.
PS
XX
XX This section represents human CASB611 cDNA. This gene exhibits
CC colon-specific expression and is highly expressed in the rectum.
CC The invention relates to human CASB cDNA sequences CASB611, CASB500,
CC CASB501, CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and
CC also to human CASB partial cDNA sequences (A72230-A72235) derived from
CC expressed sequence tags (ESTs). Expression of the human CASB genes
CC (with the exception of CASB611) is associated with colon tumours, and the
CC encoded proteins (sequences not given in the specification) represent
CC colon tumour-associated antigens. The cDNA sequences may be used in

CC diagnosing the presence or a susceptibility to a disease related to the
CC presence, expression or activity of CASB genes. Such diseases include
CC autoimmune diseases and especially colon cancer. The nucleic acid
CC sequences may also be used in genetic vaccines for the prophylaxis or
CC therapeutic treatment of colon cancer and autoimmune diseases.
XX
SQ Sequence 801 BP; 165 A; 181 C; 255 G; 200 T; 0 other;

alignment_scores:
Quality: 400.00 Length: 77
Ratio: 5.263 Gaps: 1
Percent Similarity: 98.701 Percent Identity: 97.403

alignment_block:
US-09-599-087-5 x A72224/rev ..

Align seg 1/1 to reverse of: A72224 from: 1 to: 801

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
769 ATGAGGCTTACGCTTTCAGGCTGCTGTATCTGCTGCTGCTT 720
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.AlaTppSer 33
|||||
719 CTCATCTTCTCCACAGAGGGAAGGCGCTCCTGCCACAGCGCTGCTCA 670
34 G1ArgArgThrArgLeuCysCysHisArgValProSerProAsnSerTh 50
|||||
669 GGCAGAGAGAACCAAGGCTCTGCTGCCACCGAGTCCCTAGCCCACTCAAC 620
50 rAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGlu 67
|||||
619 AAACCTGAAGAGACATCATGTGAGGCTGTAAACCATGCAACCTTGAGC 570
67 roGluProArgLeuTTPValValProGly 76
|||||
569 CAGAGCCCGCCCTTGGTGGTGGCTGGG 541

seq_name: /SID2/gcdata/geneseq/geneseqn/NA1994.DAT:Q57015

seq_documentation_block:
ID Q57015 standard; DNA; 2167 BP.
XX
XX Q57015;
AC
XX
XX 31-AUG-1994 (first entry)
DE
XX
XX PKC gamma.
DE
XX
XX 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;
KW blood vessel plaques; ss.
XX
XX Bos taurus.
OS
XX
XX WO9403609-A.
PN
XX
XX 17-FEB-1994.
PD
XX
XX 05-AUG-1993; 93WO-GB01651.
PR
XX
XX 05-AUG-1992; 92GB-001654.
PR
XX
XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA
XX
XX Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PI
XX
XX WPI; 1994-065697/08.
DR
XX
XX Eukaryotic cells transformed with mammalian phospholipid or
PT protein kinase DNA - useful in assays for compounds involved in

CC Proteins can be used in ECDN and ECDNm protein
CC Immunohistochemical assays.

XX Sequence 1979 BP; 405 A; 642 C; 602 G; 330 T; 0 other;

alignment_scores:

Quality:	85.00	Length:	67
Ratio:	2.429	Gaps:	4
Percent Similarity:	52.239	Percent Identity:	38.806

alignment_block:
US-09-599-087-5 x T27616/rev

Align seg 1/1 to reverse of: T27616 from: 1 to: 1979

```

22 ThGluGlyLysArgArgProAlaLysAla.....TTPserGlyArgar 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
983 ACTGGGGGTCTGCGCCAGGCGCGCTGACTTGGCTGCTGCGAG 934

36 gThrArgLeuCySCysHisArgVal...ProSerProAsnSerThrAsnL 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
933 AAGGAGCGTTGTTCACAGTCAGTGGCGCCGACCACTGCTGATCAT 884

52 eulysGlyHisHisValArgLeuCySLysProCySLysLeuGluPro... 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
883 TA.....GTTCTTGAGCCGCTG 867

68 .....GluProArgLeuTrpValValProGlyAlaLeuProGluIna 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
866 TTAGCTGGACACCTCGCTTCCCGGAGCCCTGCTGCTGCTGCTGAT 817

81 1 81
816 C 816

```

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:Q6134

seq_documentation_block:
ID Q6134 standard; cDNA; 2030 BP.

AC Q6134;

XX 06-OCT-1994 (first entry)

XX Human recombinant steroid hormone receptor NERI cDNA.

XX Steroid hormone receptor; NERI; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 245..1627

XX misc_binding 503..706

XX misc_binding /note= "DNA binding area of encoded protein"

XX misc_binding /tag= c

XX misc_binding /note= "ES 12 probe binding region"

XX WO9407916-A.

XX 14-APR-1994.

XX 27-SEP-1993; 93WO-US09165.

XX 07-OCT-1992; 92US-0958137.

XX (MERI) MERCK & CO INC.

XX Rodan GA, Rutledge SJ, Schmidt A, Vogel RL;

DR WPI: 1994-135511/16.

DR P-PSDB; R52980.

XX New human steroid receptor NERI and corresp. DNA - used to

XX develop assays to identify steroid hormone agonists and

XX antagonists

XX Claim 7; Page 29; 45pp; English.

XX The NERI cDNA can be used with a reporter gene (e.g. firefly

XX luciferase gene) to determine the affinity of a test sample for the

XX receptor. This can be used to identify compounds which may be

XX peroxisome proliferators, and hence, are potential hepatocarcinogens,

XX The NERI cDNA can also be used to develop novel steroid hormone

XX antagonists and agonists. This cDNA was obtained using RNA

XX prepared from osteosarcoma SAOS-2/B10 cells.

XX Sequence 2030 BP; 428 A; 655 C; 613 G; 334 T; 0 other;

Align seg 1/1 to reverse of: Q6134 from: 1 to: 2030

```

22 ThGluGlyLysArgArgProAlaLysAla.....TTPserGlyArgar 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022 ACTGGGGGTCTGCGCCAGGCGCGCTGACTTGGCTGCTGCTGCTGAG 973

36 gThrArgLeuCySCysHisArgVal...ProSerProAsnSerThrAsnL 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 AAGGAGCGTTGTTCACAGTCAGTGGCGCCGACCACTGCTGATCAT 923

52 eulysGlyHisHisValArgLeuCySLysProCySLysLeuGluPro... 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 TA.....GTTCTTGAGCCGCTG 906

68 .....GluProArgLeuTrpValValProGlyAlaLeuProGluIna 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 TTAGCTGGACACCTCGCTTCCCGGAGCCCTGCTGCTGCTGCTGAT 856

81 1 81
855 C 855

```

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:T18996

seq_documentation_block:
ID T18996 standard; DNA; 2030 BP.

XX T18996;

XX 25-SEP-1996 (first entry)

XX Human steroid receptor NER gene.

XX Steroid receptor; NER; TOFA; NGF potentiator; muscarinic agonist;

XX dopamine D1 antagonist; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 245..1630

XX /tag= a

XX WO9613519-A1.

XX 09-MAY-1996.

XX 24-OCT-1995; 95WO-US13924.
 XX
 XX 27-OCT-1994; 94US-0330283.
 XX
 XX (MEDI-) MEDICAL COLLEGE PENNSYLVANIA.
 PA (MERI) MERCK & CO INC.
 XX
 XX Friedman E, Holloway MK, Rodan GA, Rutledge SJ;
 PI Schmidt A, Vogel RL;
 XX
 XX WPI: 1996-239449/24.
 DR P-PSDB; R97982.
 XX
 XX New isolated human steroid receptor NER - used to identify cpds.
 PT which can act, e.g., as potentiators of NCF, muscarinic agonists or
 PT dopamine D1 antagonists.
 XX
 XX Claim 7; Page 57-59; 72pp; English.
 PS
 XX A gene (T18996) isolated from a human osteosarcoma SAOS-2/B10
 CC cell library codes for a novel human steroid receptor designated
 CC NER (R97982). The gene was isolated by PCR amplification using
 CC primers (T18997-99) based on consensus sequences of the DNA and
 CC ligand binding domains of a typical nuclear receptor. The gene
 CC can be incorporated into a vector, esp. pJN1ER1, and used to
 CC express NER in transfected COS cells. A chimeric gene can be
 CC created by substituting the DNA-binding domain in the NER gene
 CC with a DNA-binding domain taken from another steroid hormone
 CC receptor. Host cells expressing the chimeric gene and a reporter
 CC gene are used to identify functional ligands of the NER receptor.
 XX
 S0 Sequence 2030 BP; 428 A; 655 C; 614 G; 333 T; 0 other;

alignment_scores:
 Quality: 85.00 Length: 67
 Ratio: 2.429 Gaps: 4
 Percent Similarity: 52.239 Percent Identity: 38.806

alignment_block:
 US-09-599-087-5 x T18996/rev ..

Align seg 1/1 to reverse of: T18996 from: 1 to: 2030

```

22 ThrGluGlyLysArgArgProAlaLysAla.....TTPSerGlyArgAr 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022 ACTGGGGGCTCTGCGCCAGGCGGCGTACCTTTGGGCTGGTGGAG 973
36 gThrArgLeuCySCysHisArgVal...ProSerProAsnSerThrAsnL 52
| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 AAGGAGCGTTTGTTCGACCTGCGAGTTGGCGCCACCACTGCTGATCAT 923
52 eulysGlyHisHisValArgLeuCySLysProCysAlaLysLeuAlPro... 67
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 TA.....GTTCCTGAGCGCGTGG 906
68 .....GluProArgLeuTrpValValProGlyAlaLeuProGlyIva 81
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 TTAGCTGGACACCTTCGCTTCCCGGAGCCCTGGCTGCTGCTCAGAT 856
81 I 81
855 C 855

```

seq_name: /SIPDS2/gcdata/geneseq/geneseqn/NA1996.DAT:T30031

seq_documentation_block:

ID T30031 standard; DNA; 2030 BP.

AC T30031;

DT 19-AUG-1996 (first entry)

XX NER receptor potentiators DNA.
 DE
 XX
 XX NER receptor; potentiators; steroid hormone receptor;
 KW G-protein coupled receptor; nerve growth factor; Alzheimer disease;
 KW ocular hypertension; schizophrenia; dystonia; tardive dyskinesia;
 KW Gilles de la Tourette syndrome; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT CDS 245..1630
 FT /*Cag- a
 FT
 XX WO9613257-A1.
 PN
 XX
 XX 09-MAY-1996.
 PD
 XX
 XX 24-OCT-1995; 95WO-US13931.
 PE
 XX
 XX 27-OCT-1994; 94US-0330518.
 PR
 XX
 XX (MEDI-) MEDICAL COLLEGE PENNSYLVANIA.
 PA (MERI) MERCK & CO INC.
 XX
 XX Friedman E, Holloway MK, Rodan GA, Schmidt A, Vogel RL;
 PI
 XX WPI: 1996-239256/24.
 DR P-PSDB; R98140.
 DR
 XX
 XX Use of steroid hormone receptor NER activators - for potentiating
 PT activity of modulator of G-protein coupled receptor
 PT
 XX
 XX Disclosure; Page 49-50; 63pp; English.
 PS
 XX

CC A DNA clone (T30031) codes for a novel human steroid hormone
 CC receptor (R98140), designated NER, that is useful as a potentiator
 CC of ligands for other receptors, partic. G-protein coupled receptors.
 CC It was isolated by screening an osteosarcoma SAOS-2/B10 cDNA library
 CC using an NER fragment that was obt. by PCR amplification of SAOS-2/B10
 CC cDNA using primers (see also T30032-37) based on consensus sequences
 CC and ligand binding domains of a typical nuclear receptor. The DNA can
 CC be used for prodn. of recombinant NER using e.g. COS host cells.
 CC
 S0 Sequence 2030 BP; 428 A; 655 C; 614 G; 333 T; 0 other;

alignment_scores:
 Quality: 85.00 Length: 67
 Ratio: 2.429 Gaps: 4
 Percent Similarity: 52.239 Percent Identity: 38.806

alignment_block:
 US-09-599-087-5 x T30031/rev ..

Align seg 1/1 to reverse of: T30031 from: 1 to: 2030

```

22 ThrGluGlyLysArgArgProAlaLysAla.....TTPSerGlyArgAr 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022 ACTGGGGGCTCTGCGCCAGGCGGCGTACCTTTGGGCTGGTGGAG 973
36 gThrArgLeuCySCysHisArgVal...ProSerProAsnSerThrAsnL 52
| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 AAGGAGCGTTTGTTCGACCTGCGAGTTGGCGCCACCACTGCTGATCAT 923
52 eulysGlyHisHisValArgLeuCySLysProCysAlaLysLeuAlPro... 67
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 TA.....GTTCCTGAGCGCGTGG 906
68 .....GluProArgLeuTrpValValProGlyAlaLeuProGlyIva 81
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 TTAGCTGGACACCTTCGCTTCCCGGAGCCCTGGCTGCTGCTCAGAT 856

```


CC the course of the invention for the identification and characterisation
 CC of the polynucleotide and protein sequences.

XX Sequence 858 BP; 221 A; 244 C; 206 G; 184 T; 3 other;

alignment_scores:

Quality: 81.00 Length: 83
 Ratio: 1.884 Gaps: 5
 Percent Similarity: 51.807 Percent Identity: 32.530

alignment_block:
 US-09-599-087-5 x F18164/rev ..

Align seg 1/1 to reverse of: F18164 from: 1 to: 858

```

3 LeuLeuValLeuSerLeuLeuCysIleLeuLeuLeuLeuPheSerIle 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
234 CTCTGCTCTGCGAGTCGCTTCACTCTCTCCACCTCTGCCA..... 191
19 ephSerThrIle.....GlyArgArg..... 27
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 .FTTCGACCATCGGTAGAGCGCGAGAAATACATATCAATAAGT 142
28 .....ProAlaIysAlaTrpSerGly 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 GCTTGGCAGAGCCTGCGACAACTGAGCTCCGGGAAAGCTGGCGGGT 92
35 ArgArgThrArgLeuCysSHSArgValProSerProAsnSerThrAs 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
91 .....TGCCATCCGCACACCAAGGCCCAAAACTCTAG 60
51 nLeuIysGly.....HisIysValArgLeuCysLys 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
59 GCTGACAGGCGCGCCGCCACCCCTCCGCATCATCTCTCTGGCGAG 11

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:C74372

seq_documentation_block:

ID C74372 standard; cDNA; 2206 BP.

XX C74372;

XX 02-FEB-2001 (first entry)

XX Human secreted protein gene 36 SEQ ID NO:46.

XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antinflammatory; antitumor; vulnerary; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
 KW food additive; preservative; chromosome identification; cancer;
 KW female reproductive system disorder; immune disorder; wound healing;
 KW cardiovascular disorder; neurological disease; infectious disease;
 KW infection; ss.

XX Homo sapiens.

XX WO200058340-A2.

XX 05-OCT-2000.

XX 23-MAR-2000; 2000WO-US07724.

XX 26-MAR-1999; 99US-0126510.

XX 07-JAN-2000; 2000US-0174850.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, KomatsuIis G;

XX WPI: 2000-594638/56.

XX P-PSDB; B39437.

XX Filly nucleic acid molecules encoding human secreted proteins, useful
 PT in the prevention, treatment and diagnosis of cancer, immune disorders,
 PT cardiovascular disorders and neurological diseases -
 XX
 XX
 PS Claim 1; Page 347-348; 391pp; English.

CC The polynucleotide sequences given in C74337 to C74386 encode the human
 CC secreted proteins given in B39402 to B39451. B39452 to B39484 represent
 CC human secreted polypeptide sequences and proteins homologous to them,
 CC which are given in the exemplification of the present invention. Human
 CC secreted proteins have activities based on the tissues and cells the
 CC genes are expressed in. Example of activities include: cytostatic;
 CC immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
 CC hepatotropic; antidiabetic; antinflammatory; antitumor; vulnerary;
 CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
 CC The polynucleotides and polypeptides are useful for preventing, treating
 CC or ameliorating a medical condition in e.g. humans, mice, rabbits, goats,
 CC horses, cats, dogs, chickens or sheep. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities. The polynucleotide are useful for chromosome
 CC identification. They are also useful as probes for diagnosing a disorder
 CC related to the female reproductive system, particularly breast and/or
 CC ovary cancer. They are also useful in the gene therapy of breast and
 CC ovarian cancer. Secreted protein nucleic acids, proteins, antibodies,
 CC agonists and antagonists are useful in the diagnosis, treatment and
 CC prevention of: (a) cancer; (b) immune disorders; (c) cardiovascular
 CC disorders; (d) wound healing; (e) neurological diseases; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. C74328 to C74336 and B39401 represent sequences used in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 2206 BP; 528 A; 456 C; 500 G; 691 T; 31 other;

alignment_scores:

Quality: 79.50 Length: 91
 Ratio: 1.656 Gaps: 6
 Percent Similarity: 52.747 Percent Identity: 32.967

alignment_block:

US-09-599-087-5 x C74372 ..

Align seg 1/1 to: C74372 from: 1 to: 2206

```

7 SerSerLeuLeuCysIleLeuLeuLeuCysPheSerThrIle 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1012 GCAGCTGCTGTTGTTTGTGTTGTTGTTGTTGTTGTTGAGACACA 1061
23 nGlyIysArg. ArgProAlaIysAlaTrpSerGly.....ArgArgTh 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1062 GCTCGACACTGTCGCCCA.....AGCTGAGTGCAGTGTGCGATCTCA 1105
37 rArgLeuCysCysSHSArgVal.....ProS 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1106 GCAGCTGCTGTTTAAAGAGTCTTCTACTCTTCTTCAATCCAG 1155
46 eProAsnSerThrAsnLeuIysGlyHisValArgLeuCysLysPro 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1156 ACCCTGTAAATTCCTCCACTTCTCTGCGC..... 1181
63 CysLysLeuGluProGlu.....ProArgLeuTrpValVa 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1182 TGCCCCCTCCCAAGAAAGCGCCCTCTGCTCCCTCCACATGGGTGT 1231
74 lProGlyAlaLeuProGlnVal 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1232 GCCCGAGTTCAGGAGCAGAGTCT 1253

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:Q59506

seq_documentation_block:

ID Q59506 standard; cDNA; 433 BP.

```

XX AC Q59506;
XX DT 16-MAR-1994 (first entry)
XX DE Human brain Expressed Sequence Tag EST00507.
XX DE Gene transcription product; genetic markers; tagging; in vivo;
XX KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX OS Homo sapiens.
XX PN WO9316178-A.
XX PD 19-AUG-1993.
XX PF 12-FEB-1993; 93WO-US01294.
XX PR 12-FEB-1992; 92US-0837195.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX PI Adams MD, Moreno RF, Venter CJ;
XX DR MPI; 1993-272882/34.
XX PS
XX PT Enriched oligonucleotides and corresp. sequences - used as
XX PT markers for human genes transcribed in-vivo, facilitate tagging
XX PT of most human genes
XX PS
XX PS Example 4; Page 186; 500pp; English.
XX CC The Expressed Sequence Tag was isolated from a human brain CDNA
XX CC library as part of a large set of ESTs which can be used as markers
XX CC for human genes transcribed in vivo. They can be used to facilitate
XX CC tagging of most human genes, for mapping locations of expressed genes
XX CC on chromosomes, for individual or forensic identification, for mapping
XX CC locations of disease-associated genes, for identification of tissue
XX CC type, and for prep. of antisense sequences, probes and constructs.
XX CC EST00507 has a "poor" coding probability as evaluated using the
XX CC coding-region prediction program CRM. See also Q59041-Q61440.
XX CC
XX SO Sequence 433 BP; 114 A; 88 C; 85 G; 142 T; 4 other;

Alignment_scores:
Quality: 79.00 Length: 42
Ratio: 3.950 Gaps: 1
Percent Similarity: 47.619 Percent Identity: 35.714

alignment_block:
US-09-599-087-5 x Q59506 ..
Align seg 1/1 to: Q59506 from: 1 to: 433

32 TTPSeGlyArgArgThrArgLeuGlyCysHisArgValProSerProAs 48
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 TGGACAGACACCGACACAGAAAGCTGC..... 92
48 nserThrAnLeuLysGlyHisHisValArgLeuGlyCysLysProCysAlaL 65
93 .....CTCAGTTTGCTGCTCCCTCCGCAAG 117
65 eugluProgluProalrGleuTrpVal 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 CAGAGCGCTGAGACAGAGATTGGGTA 143

seq_name: /SID52/gcgdelta/geneseq/geneseqn/NA1997.DAT:T77840
seq_documentation_block:
ID T77840 standard; mRNA; 1735 BP.
XX AC T77840;

```

```

XX AC 17-MAR-1998 (first entry)
XX DT Human melanoma associated delayed early response variant mRNA sequence.
XX DE Melanoma associated delayed early response gene; MADER gene;
XX DE MADER protein; growth alteration; malignant melanoma; breast carcinoma;
XX KW cancerous condition; MADER translocation event; MADER immunogen;
XX KW MADER antigen; ss.
XX OS Homo sapiens.
XX PN WO9728193-A1.
XX PD 07-AUG-1997.
XX PF 30-JAN-1997; 97WO-US01586.
XX PR 30-JAN-1997; 97US-0593563.
XX PR 30-JAN-1996; 96US-0593563.
XX PA (MELC-) MELCORP DIAGNOSTICS INC.
XX PI Johnson JP;
XX DR MPI; 1997-402557/37.
XX DR P-PSDB; W24230.
XX PS
XX PT Monoclonal antibody which binds to the MADER protein - used for
XX PT detecting cancerous conditions, such as breast carcinoma
XX PT
XX PS Disclosure: Fig 3; 69pp; English.
XX CC The present sequence represents a melanoma associated delayed early
XX CC response (MADER) gene, which encodes a novel 55 kDa nuclear protein. This
XX CC protein is associated with growth alterations in malignant melanomas and
XX CC other cancerous conditions, and is over-expressed in human malignant
XX CC melanomas. Several variants of the protein have been identified
XX CC (W24228-31), the present sequence encoding a variant which binds
XX CC erg-1 and inhibits its activity. Chromosomal rearrangement of MADER can
XX CC be detected by hybridizing immobilised chromosomal target DNA, that has
XX CC been rendered single stranded and is obtained from a cell suspected of
XX CC having undergone a MADER translocation event with a single stranded
XX CC oligonucleotide probe complementary to a MADER nucleotide sequence. The
XX CC probe contains a moiety capable of direct or indirect visualisation.
XX CC Antibodies raised against the MADER protein can be used for detecting a
XX CC cancerous condition, particularly melanoma malignancies, and especially
XX CC a breast carcinoma. A composition comprising a MADER immunogen and a
XX CC pharmaceutically acceptable vehicle can be used to elicit an immune
XX CC response against a cell which over-expresses a MADER antigen.
XX CC
XX SO Sequence 1735 BP; 357 A; 516 C; 594 G; 268 T; 0 other;

Alignment_scores:
Quality: 79.00 Length: 67
Ratio: 2.257 Gaps: 4
Percent Similarity: 52.239 Percent Identity: 31.343

alignment_block:
US-09-599-087-5 x T77840/rev ..
Align seg 1/1 to reverse of: T77840 from: 1 to: 1735

22 ThrGluGlyLysArgArgProAlaLysAlaTrpSerGlyArgArgThr.. 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1602 ACACAGCGGCTGAGCGCGCGCGCTGCGTGGAGACAGCGCGGCGAG 1553
38 .....ArgLeuGlyCysHis.....ArgValProAs 46

```


34 l y n g i n l a g l e u s c y s t i s a l g v a l p r o s e t p r o a s n s e t i n r 50

215 GCAGGAGAACCGAGCTGCTGCTCCACGAGTCCCTAGCCCACTCAACA 264

51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67
|||||
265 AACCTGAAGACATCATGTGAGGCTCTGTAAACATGACAGCTTGAGCC 314
|||||
67 OGUPrArgLeuTrpValValProGlyAlaLeuProGlyVal 81
|||||
315 AGAGCCCCCGCTTTGGGTGGTGGCTGGGACATCCACAGGTG 357
|||||

seq_name: gb_pat1:AX027773

seq_documentation_block: 797 bp DNA PAT 16-SEP-2000
LOCUS AX027773
DEFINITION Sequence 7 from Patent W00043509.
ACCESSION AX027773
VERSION AX027773.1 GI:10188625
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS Vinals-Bassols,C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 7 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
source 1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 164 a 180 c 254 g 199 t
ORIGIN

alignment_scores:
Quality: 400.00 Length: 77
Ratio: 5.263 Gaps: 1
Percent Similarity: 98.701 Percent Identity: 97.403

alignment_block:
US-09-599-087-5 x AX027773/rev ..

Align seg 1/1 to reverse of: AX027773 from: 1 to: 797

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
765 ATGAGGCTTCTAGCTCTTCACAGCTGCTGTATCCTGCTCTGCTT 716
|||||
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.Alatpser 33
|||||
715 CTCATCTTCTCCACAGAGGAGGAGGCGCTCGCCACAGGCGCTGCTCA 666
|||||
34 GAlArGrThrArgLeuCysCysHisArgValProSerProAsnSerTh 50
|||||
665 GCGAGGAGAACGAGGCTCTCTCCACGAGTCCCTAGCCCACTCAAC 616
|||||
50 rAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67
|||||
615 AAACCTGAAGACATCATGTGAGGCTCTGTAAACATGACAGCTTGAGC 566
|||||
67 rOGUPrArgLeuTrpValValProGly 76
|||||
565 CAGAGCCCCCGCTTTGGGTGGTGGCTGGGACATCCACAGGTG 537
|||||

seq_name: gb_pat1:AX027767

seq_documentation_block: 801 bp DNA PAT 16-SEP-2000
LOCUS AX027767
DEFINITION Sequence 1 from Patent W00043509.
ACCESSION AX027767
VERSION AX027767.1 GI:10188619
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 801)
AUTHORS Vinals-Bassols,C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 1 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
source 1..801
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 165 a 181 c 255 g 200 t
ORIGIN

alignment_scores:
Quality: 400.00 Length: 77
Ratio: 5.263 Gaps: 1
Percent Similarity: 98.701 Percent Identity: 97.403

alignment_block:
US-09-599-087-5 x AX027767/rev ..

Align seg 1/1 to reverse of: AX027767 from: 1 to: 801

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
769 ATGAGGCTTCTAGCTCTTCACAGCTGCTGTATCCTGCTCTGCTT 720
|||||
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.Alatpser 33
|||||
719 CTCATCTTCTCCACAGAGGAGGAGGCGCTCGCCACAGGCGCTGCTCA 670
|||||
34 GAlArGrThrArgLeuCysCysHisArgValProSerProAsnSerTh 50
|||||
669 GCGAGGAGAACGAGGCTCTCTCCACGAGTCCCTAGCCCACTCAAC 620
|||||
50 rAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67
|||||
619 AAACCTGAAGACATCATGTGAGGCTCTGTAAACATGACAGCTTGAGC 570
|||||
67 rOGUPrArgLeuTrpValValProGly 76
|||||
569 CAGAGCCCCCGCTTTGGGTGGTGGCTGGG 541
|||||

seq_name: gb_tol:AF152002

seq_documentation_block: 744 bp mRNA ROD 29-JUN-1999
LOCUS AF152002
DEFINITION Rattus norvegicus unknown mRNA sequence.
ACCESSION AF152002
VERSION AF152002.1 GI:5257466
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 744)
AUTHORS Rossmann,T.G. and Li,P.
TITLE 2C9-like sequence expressed in lead-resistant rat glioma cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 744)
AUTHORS Rossmann,T.G. and Li,P.
TITLE Direct Submission.
JOURNAL Submitted (12-MAY-1999) Environmental Medicine, NYU School of
Medicine, 57 Old Forge, Tuxedo, NY 10987, USA
FEATURES
source 1..744
/organism="Rattus norvegicus"
/db_xref="taxon:10116"


```

/cell_line="Pbri"
/tissue_type="glioma"
/notes="Lead resistant cells derived from C6 cell line;
Dolzanskaya et al., 1998, Biological Trace Element Res.
65:31-43"
1..744
/note="sequence expressed in Pbri cells; not expressed in
C6 cells; sequence resembles 2c9, found in cells
overexpressing fos; Pbri cells do not overexpress fos"
BASE COUNT      208 a      216 c      146 g      174 t
ORIGIN

misc_feature

alignment_scores:
      Quality: 183.00      Length: 81
      Ratio: 3.102      Gaps: 2
      Percent Similarity: 72.840      Percent Identity: 46.914

alignment_block:
US-09-599-087-5 x AF152002 ..

Align seg 1/1 to: AF152002 from: 1 to: 744

1 MetATGLeuValLeuSerSerLeuLeuCyAlleuLeuLeuCySph 17
|||||.....:|||||:..:|||||:
53 ATCGAGCTTCTCACCCCTCCGGTGTGTTCTTCACTGCTTCCTCT 102
17 eSerIlePheSerThrGluGluLysAspArgProAlaLysAlaTrpSerG 34
:.....:|||||:.....:|||||
103 CTCGGTCTCTCTCCAGAGGAGAAAGCGCTCCGCCAAGTTC..... 145
34 LysATGTrpThrArgLeuCySuaHisArgValProSerProAsnSerThr 50
:::..|||.....:|||||
146 ..CCGAAACATCAGACCCCGCGCTGCATCTATCTCCTAGATCAACACATA 193
51 AsnLeuLysGlnHisValArgLeuCyAlaProCysLysLeuGluTrp 67
:::|||||:.....:|||||:..:
194 ACCTGGAAAGCAACACACAGACCTCGACGACATGCAGAGA...AAGCT 240
67 oguProArgLeuTrpValProGluAlaLeuProGlnVal 81
|||:::|||||.....:|||||:
241 AGAATCCAAATTCATGGGTGGTCCCTGGGGCTCCACACAGATA 283

seq_name: gb_htg8:AC022389

seq documentation_block:
LOCUS      AC022389      258746 bp      DNA
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT
SEQUENCE, 56 unordered pieces.
ACCESSION      AC022389
VERSION      AC022389.3 GI:7677724
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 258746)
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 258746)
Smith,D.R.
Direct Submisson
Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 10, 2000 this sequence version replaced gi:7209016.
COMMENT
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

Project Information	
Center project name: hg137	
----- Summary Statistics -----	
Sequencing vector:	N/A
Chemistry:	Dye-terminator Big Dye; 100% of reads
Assembly program:	Phrap; version 990315
Consensus quality:	201841 bases at least Q40
Consensus quality:	230471 bases at least Q30
Consensus quality:	236618 bases at least Q20
Insert size:	258746; sum-of-contigs
Quality coverage:	5.4x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
1	1122: contig of 1122 bp in length
*	gap of unknown length
*	1123: contig of 1119 bp in length
*	gap of unknown length
*	2242: contig of 1293 bp in length
*	gap of unknown length
*	3535: contig of 1347 bp in length
*	4882: contig of 1163 bp in length
*	6044: contig of 1050 bp in length
*	7094: contig of 1050 bp in length
*	6045: contig of 1400 bp in length
*	7095: contig of 1245 bp in length
*	8495: contig of 1245 bp in length
*	9740: contig of 1344 bp in length
*	11083: contig of 1145 bp in length
*	11084: contig of 1049 bp in length
*	12229: contig of 1470 bp in length
*	13278: contig of 1073 bp in length
*	14748: contig of 1013 bp in length
*	15821: contig of 1095 bp in length
*	16834: contig of 1130 bp in length
*	17929: contig of 1062 bp in length
*	19059: contig of 1074 bp in length
*	20121: contig of 1047 bp in length
*	21195: contig of 1449 bp in length
*	22242: contig of 1306 bp in length
*	23691: contig of 1216 bp in length
*	24997: contig of 1381 bp in length
*	26213: contig of 1677 bp in length
*	27594: contig of 1984 bp in length
*	29271: contig of 1880 bp in length
*	31254: contig of 1542 bp in length
*	33135: contig of 1542 bp in length


```

192  ..AACTGGAAAGGAACACACAGACCCCTGCAGACCATCAAA...AAG 236
67  ProgluPrOArGlLeuTrP_ValValProGlLyAlaLeuProGlInVal 81
193  CTGAGAAATCCACATTCATGGGCGTGGCTGGGCGCTCCACAGATA 282
seq_name: gpl_hcgs:AC022389

seq_documentation_block:
LOCUS      AC022389      258746 bp      DNA
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5. WORKING DRAFT
SEQUENCE, 56 unordered pieces.
AC022389
AC022389.3 GI:7767724
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 258746)
AUTHORS   Smith,D.R.
TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
          Sequence Data
          Unpublished
          2 (bases 1 to 258746)
          Smith,D.R.
          Direct Submission
          Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
          Street, Waltham, MA 02453, USA
          On May 10, 2000 this sequence version replaced gi:7209016.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1122: contig of 1122 bp in length
*
* 1123 2241: gap of unknown length
*         contig of 1119 bp in length
*
* 2242 3534: gap of unknown length
*         contig of 1293 bp in length
*
* 3535 4881: gap of unknown length
*         contig of 1347 bp in length
*
* 4882 6044: gap of unknown length
*         contig of 1163 bp in length
*
* 6045 7094: gap of unknown length
*         contig of 1050 bp in length
*
* 7095 8494: gap of unknown length
*         contig of 1400 bp in length
*
* 8495 9739: gap of unknown length
*         contig of 1245 bp in length

```

*	9740	11083:	contig of 1344 bp in length
*			gap of unknown length
*	11084	12228:	contig of 1145 bp in length
*			gap of unknown length
*	12229	13277:	contig of 1049 bp in length
*			gap of unknown length
*	13278	14747:	contig of 1470 bp in length
*			gap of unknown length
*	14748	15820:	contig of 1073 bp in length
*			gap of unknown length
*	15821	16833:	contig of 1013 bp in length
*			gap of unknown length
*	16834	17928:	contig of 1095 bp in length
*			gap of unknown length
*	17929	19058:	contig of 1130 bp in length
*			gap of unknown length
*	19059	20120:	contig of 1062 bp in length
*			gap of unknown length
*	20121	21194:	contig of 1074 bp in length
*			gap of unknown length
*	21195	22241:	contig of 1047 bp in length
*			gap of unknown length
*	22242	23590:	contig of 1449 bp in length
*			gap of unknown length
*	23591	24996:	contig of 1306 bp in length
*			gap of unknown length
*	24997	26212:	contig of 1216 bp in length
*			gap of unknown length
*	26213	27593:	contig of 1381 bp in length
*			gap of unknown length
*	27594	29270:	contig of 1677 bp in length
*			gap of unknown length
*	29271	31254:	contig of 1984 bp in length
*			gap of unknown length
*	31255	33134:	contig of 1880 bp in length
*			gap of unknown length
*	33135	34676:	contig of 1542 bp in length
*			gap of unknown length
*	34677	36624:	contig of 1948 bp in length
*			gap of unknown length
*	36625	38260:	contig of 1636 bp in length
*			gap of unknown length
*	38261	39573:	contig of 1313 bp in length
*			gap of unknown length
*	39574	42491:	contig of 2918 bp in length
*			gap of unknown length
*	42492	44172:	contig of 1681 bp in length
*			gap of unknown length
*	44173	45456:	contig of 1284 bp in length
*			gap of unknown length
*	45457	47973:	contig of 2517 bp in length
*			gap of unknown length
*	47974	49274:	contig of 1301 bp in length
*			gap of unknown length
*	49275	51062:	contig of 1788 bp in length
*			gap of unknown length
*	51063	53455:	contig of 2393 bp in length
*			gap of unknown length
*	53456	55778:	contig of 2323 bp in length
*			gap of unknown length
*	55779	57022:	contig of 1244 bp in length
*			gap of unknown length
*	57023	59210:	contig of 2188 bp in length
*			gap of unknown length
*	59211	60883:	contig of 1673 bp in length
*			gap of unknown length
*	60884	62917:	contig of 2034 bp in length
*			gap of unknown length
*	62918	65997:	contig of 3080 bp in length
*			gap of unknown length
*	65998	68398:	contig of 2401 bp in length
*			gap of unknown length
*	68399	74552:	contig of 6154 bp in length

```

* 74553 78738: contig of 4186 bp in length
* 78739 83546: contig of 4808 bp in length
* 83547 90669: contig of 7123 bp in length
* 90670 100398: contig of 9729 bp in length
* 100399 108591: contig of 8193 bp in length
* 108592 121396: contig of 12805 bp in length
* 121397 131100: contig of 9704 bp in length
* 131101 148369: contig of 17269 bp in length
* 148370 168804: contig of 20435 bp in length
* 168805 210111: contig of 41307 bp in length
* 210112 258746: contig of 48635 bp in length.
Location/Qualifiers
* 1. 258746
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="CHROM 10"
  /clone="RP11-124L5"
  /clone_lib="RPC1-11"

```

```

BASE COUNT 73754 a 58179 c 55947 g 70833 t 33 others
ORIGIN

```

```

alignment_scores:
  Quality: 136.50      Length: 109
  Ratio: 2.730         Gaps: 4
  Percent Similarity: 45.872   Percent Identity: 39.450

```

```

alignment_block:
  US-09-599-087-5 x AC022389

```

```

Align seg 1/1 to: AC022389 from: 1 to: 258746

```

```

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuCysP 17
|||||
108257 ATGAGCTTCTAGTCTTCCAGCCTGCTGTATCTCTGCT 108306
|||||

```

```

17 eSerIlePheSerThrIleuGlyIysArg..... 26
|||||
108307 CTCATCTTCTCCACAGAGTAGGCGACGCCAGGATCCCTG 108356
|||||

```

```

26 ..... 26
108357 ACAGAGATTTCAGCATCTGGGAAGACTGTGATGAGATTGTTGGAGGC 108406
|||||

```

```

27 ArgProAlaIlysalatrpserglyararararararar 43
|||||
108407 AGGCTTGGCGGCGGCTTGGCAGCAGCCCGCCTC..... 108445
|||||

```

```

43 gValProSerProAsnSerThrAsnLeuIysGlyIshIshIshValArgLeuc 60
|||||
108446 ....CCGTC.....CTCCGGGCTC 108461
|||||

```

```

60 yAlspProcysIysLeuGluProGluProArgLeuTrpValVal..... 74
|||||
108462 TCCTCCCGCCCTCTCTGCGGCTGCGCGCGCTCTCTCGGCGGCG 108511
|||||

```

```

75 .....ProGlyAlaLeuPro 79
|||||
108512 GGCCCGCGCTCTCCCTGGGACCGCC 108538
|||||

```

```

seq_name: gb_htg20:AL158017

```

```

seq_documentation_block:
LOCUS      AL158017 118429 bp  DNA           23-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP5-893J12 map q32.1-32.3, ***
ACCESSION  AL158017
VERSION    AL158017.4 GI:9926467
KEYWORDS   HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 118429)
AUTHORS   Plumb,B.
TITLE      Direct Submission
JOURNAL    Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            On Aug 25, 2000 this sequence version replaced gi:9212511.
COMMENT    ----- genome center
            Center: Sanger Centre
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            ----- Project Information
            Center project name: d9893J12
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 117880 bases at least Q40
            Consensus quality: 118051 bases at least Q40
            Consensus quality: 118225 bases at least Q20
            Insert size: 118329; sum-of-contigs
            Insert size: 130142; 2.1% error; agarose-fp
            Quality coverage: 6.60x in Q20 bases; sum-of-contigs
            Quality coverage: 6.06x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1 22002: contig of 22002 bp in length
            * 22003 22102: gap of 100 bp
            * 22103 118429: contig of 96327 bp in length.
            Location/Qualifiers
            1. 118429
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="1"
              /map="q32.1-32.3"
              /clone="RP5-893J12"
              /clone_lib="RPC1-5"
              /note="assembly_fragment:00390"
            1. 22002
              /note="assembly_fragment:00587"
            misc_feature
              22103..118429
              vector_side:left
            misc_feature
              22103..118429
              /note="assembly_fragment:00587"
            BASE COUNT 32098 a 28666 c 27720 g 29844 t 101 others
            ORIGIN

```

```

alignment_scores:
  Quality: 87.50      Length: 89
  Ratio: 1.902         Gaps: 7
  Percent Similarity: 51.685   Percent Identity: 38.202

```

```

alignment_block:
  US-09-599-087-5 x AL158017

```

FEATURES	source
* be preserved.	
* 1	
* 7575 7674: gap of 100 bp	
* 7675 20397: contig of 12733 bp in length	
* 20398 20497: gap of 100 bp	
* 20498 25271: contig of 4774 bp in length	
* 25272 25371: gap of 100 bp	
* 25372 42498: contig of 17127 bp in length	
* 42499 42598: gap of 100 bp	
* 42599 70455: contig of 27857 bp in length	
* 70456 70555: gap of 100 bp	
* 70556 82770: contig of 12215 bp in length	
* 82771 82870: gap of 100 bp	
* 82871 87579: contig of 4709 bp in length	
* 87580 87679: gap of 100 bp	
* 87680 118093: contig of 30414 bp in length	
* 118094 118193: gap of 100 bp	
* 118194 124413: contig of 6220 bp in length	
* 124414 124513: gap of 100 bp	
* 124514 127471: contig of 2958 bp in length	
* 127472 127571: gap of 100 bp	
* 127572 132634: contig of 5063 bp in length	
* 132635 133734: gap of 100 bp	
* 133735 140792: contig of 8058 bp in length	
* 140793 140892: gap of 100 bp	
* 140893 157198: contig of 16306 bp in length	
* 157199 157298: gap of 100 bp	
* 157299 179799: contig of 20661 bp in length	
* 179800 178079: gap of 100 bp	
* 178080 194041: contig of 15962 bp in length	
* 194042 194141: gap of 100 bp	
* 194142 205069: contig of 10928 bp in length	
* 205070 205169: gap of 100 bp	
* 205170 209120: contig of 3951 bp in length	
* 209121 209220: gap of 100 bp	
* 209221 212390: contig of 3170 bp in length.	
Location/Qualifiers	
1. 212390	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="1"	
/clone="RP1-328D5"	
/clone_1lb="RPc1-11.2"	
1. 7574	
/note="assembly-fragment:02000	
fragment_chain:1	
clone_end:17	
vector_side:left	
7675..20397	
/note="assembly-fragment:02052	
fragment_chain:1"	
20498..25271	
/note="assembly-fragment:02543	
fragment_chain:1"	
25372..42498	
/note="assembly-fragment:02521	
fragment_chain:1"	
42599..70455	
/note="assembly-fragment:02376	
fragment_chain:1"	
70556..82770	
/note="assembly-fragment:01906	
fragment_chain:2"	
82871..87579	
/note="assembly-fragment:02522	
fragment_chain:2"	
87680..118093	
/note="assembly-fragment:00955	
fragment_chain:2"	
118194..124413	
/note="assembly-fragment:01353	
fragment_chain:2"	
124514..127471	
124514..127471	

```

misc_feature /note="assembly_fragment:01329
fragment_chain:2"
misc_feature /note="assembly_fragment:02593
fragment_chain:2"
misc_feature /note="assembly_fragment:02846
fragment_chain:3"
misc_feature /note="assembly_fragment:01653
fragment_chain:3"
misc_feature /note="assembly_fragment:00126
fragment_chain:3"
misc_feature /note="assembly_fragment:02035
fragment_chain:3"
misc_feature /note="assembly_fragment:02081
fragment_chain:3"
misc_feature /note="assembly_fragment:02342
fragment_chain:3"
misc_feature /note="assembly_fragment:01418
clone_end:SP6
vector_side:right"
BASE COUNT 58812 a 47762 c 47014 g 57095 t 1707 others
ORIGIN

alignment_scores:
Quality: 87.50 Length: 89
Ratio: 1.902 Gaps: 7
Percent Similarity: 51.685 Percent Identity: 38.202

alignment_block:
US-09-599-087-5 x AL356275/rev ..
Align seg 1/1 to reverse of: AL356275 from: 1 to: 212390

4 LeuValLeuSerSerLeuLeu.....CysIleLeuLeuLeuLeuLeu 17
||||| ||||| ||||| ||||| |||||
121153 CTCCTTTTAAGGAGCAATTAAGTGTGCGTGCATGTGCAATTAAGCCT 121104
||||| ||||| ||||| ||||| |||||
17 e.....SerIlePheSerThrGluGlyLysArgArgProAlaLysAla 32
:||||| ||||| ||||| ||||| |||||
121103 ATTAAGACGTGCTATCTCTACAGGGCAAAACCTCTGCCCTCTGCTCT 121054
||||| ||||| ||||| ||||| |||||
32 rPserGly.....ArgArgThrArgLeuLeu 40
||||| ||||| ||||| ||||| |||||
121053 GGAGGGGCAAACTACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 121004
||||| ||||| ||||| ||||| |||||
41 CysHisArgValProSerProAsnSer.....ThrAsnLeuLeuG 54
||||| ||||| ||||| ||||| |||||
121003 TGTCAACCGATATCGAGACCCAGCTTCTGCGAGCCCAACTCCACCG 120954
||||| ||||| ||||| ||||| |||||
54 lYHisHis...ValArgLeuLeuLysProCysLysLeuGluProGlu 69
||||| ||||| ||||| ||||| |||||
120953 GGAGAGCAAGAGTCATGTGAGTGAGAGCCCTGCGTCTGTATAGCTCTCT 120904
||||| ||||| ||||| ||||| |||||
69 oArgLeuTrpVal 73
||||| ||||| ||||| ||||| |||||
120903 CCGCCCTCGTGTG 120891

seq_name: gb_pat1:A37236

seq_documentation_block:
LOCUS A37236 2167 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 5 from Patent WO9403609.
ACCESSION A37236
VERSION A37236.1 GI:2294349
KEYWORDS

```

```

SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 2167)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EDUCAROTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 5 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
Source Location/Qualifiers
1..2167
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 428 a 700 c 611 g 428 t
ORIGIN

alignment_scores:
Quality: 86.50 Length: 83
Ratio: 2.471 Gaps: 4
Percent Similarity: 42.169 Percent Identity: 31.325

alignment_block:
US-09-599-087-5 x A37236 ..
Align seg 1/1 to: A37236 from: 1 to: 2167

18 SerIlePheSerThrGluGlyLysArgArg.....ProAlaLysAl 31
||||| ||||| ||||| ||||| |||||
1882 TCGACAGATCTTCACACTCGGGGGGCGCGCGCTGACACCCCTGACCC 1931
||||| ||||| ||||| ||||| |||||
31 aTrp.....SerGlyArg 36
||||| ||||| ||||| ||||| |||||
1932 CTGGTCTGCGCAGCATGCAGCGCTGAGTTCAGAGCTTCACTATGCT 1981
||||| ||||| ||||| ||||| |||||
36 rGThrArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 52
||||| ||||| ||||| ||||| |||||
1982 CAACCGGATTCGTGTCACCGGATGCGCGGAGCCGACCATCAGCCCA... 2028
||||| ||||| ||||| ||||| |||||
2029 .....CGCCTGTGCGCAGTCACTGATATCCACCTGCCGCCA 2063
||||| ||||| ||||| ||||| |||||
64 .....LysLeuGluProGluProArgLeuTrpValPro 75
||||| ||||| ||||| ||||| |||||
2064 CCAGCGCTCCCAAGCGCTCTCTCTCCGCCCGGCTTGGCCCTGCT 2112

seq_name: gb_cm:BOVPRG

seq_documentation_block:
LOCUS BOVPRG 2168 bp mRNA MAN 27-APR-1993
DEFINITION Bovine gamma type protein kinase C mRNA.
ACCESSION M13976
VERSION M13976.1 GI:163525
KEYWORDS protein kinase C-gamma.
SOURCE Bovine brain, cDNA to mRNA, clones lambda-bPKC-gamma[19,551].
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 2168)
Cousens, L., Parker, P.J., Rhee, L., Yang-Feng, T.L., Chen, E.,
Waterfield, M.D., Franke, U. and Ullrich, A.
Multiple, distinct forms of bovine and human protein kinase C
suggest diversity in cellular signaling pathways
Science 233, 859-866 (1986)
86289426
JOURNAL Science 233, 859-866 (1986)
MEDLINE 86289426
FEATURES
Source Location/Qualifiers
1..2168
/organism="Bos taurus"
/db_xref="taxon:9913"
<1..2049
CDS

```

```
/note="gamma type protein kinase C"
/codon_start=1
/protein_id="AAA30704.1"
/db_xref="gi:163526"
/translation="RPLECRKALRQKVVHKKFAPFKOPTFCSHCPTFMGI
KGQOCOCVCSFVVRRCHEVTEFCPGAGKOPDDPRNKHKEFLHSYSPFDHC
GSLIYGLVHOGKCSCEMNVHRCVRPSLIGVDHREGRLOLETRAPSPDIHY
TVGERANLIPMDPNGLSDPYVKIKLIPPRNLTOKRTVTATLNPVNETFVNLKP
GDERLSVEVMDRISNDPMGAMSGVSELPAVDGWTKLNDGEGEYVNPVA
DADCNLLOKEACNYPLELERYRTGPSSPISPSPPDSKRCFPGASGRHLIS
DSEFLMVLGKSGFVGLAERGSDELAIKIKDVIYODDVCCTLKERVLHIS
RGGRAPHEFLTOLSTFOTPDRLYFVMEYVGGDLAMHIQODGKEKHAFAVEIA
IGLEFLHMOGITLTDLKDNDYMEGHKIKTDPMCKENPFGSTTRFCTGPDYIA
PEITAYOPKGSVDMWSRCVLLYLELACGPPDGDDELEQALHEQVTVTKLSRE
AVAIKGFLLNHPARKRLSGPDGETIRAHGFRRIMDWRLERLEIAPFRPCGRS
GENDFKEFTRAPALTPPDRLVLASIDAEFOGFYVNPDEVHPDARSPISPTPVPM"
```

BASE COUNT 428 a 701 c 611 g 428 t
ORIGIN Unreported.

alignment_scores:
Quality: 86.50 Length: 83
Ratio: 2.471 Gaps: 4
Percent Similarity: 42.169 Percent Identity: 31.325

alignment_block:
US-09-599-087-5 x BOVPGK ..

Align seg 1/1 to: BOVPGK from: 1 to: 2168

```
18 SerIlePheSerThrgluGlyLysArgArg.....ProAlaLysAl 31
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1883 TCACACAAGTTCTTCACTCGGCGCGCGCGGTACACCCCTGACCGC 1932
31 ATP.....SerGlyArgA 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1933 CTGGCTTCGGCCAGCATCGACGAGGCTGAGCTTCACGCTTCACTATGT 1982
36 rGthrArgLeuGlyCysHisArgValProSerProAsnSerThrasLeu 52
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1983 CAACCCGGAATTGCTGACCCGCGATGCGCCAGCCCATGACCCCA... 2029
53 LysGlyHisHisValArgLeuGlyCysLysProGly..... 63
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2030 .....CGCGTGTCCAGTCAATGATCCACGCTGCGGCCA 2064
64 .....LysLeuGluProGluProArgLeuTyrValPro 75
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2065 CCAGGCGGCCCAACGAGCTCCCTCCCTCCGCGCGCTTGGCCCTGCCCT 2113
```

seq_name: gb_hcgl6:AC073252

seq_documentation_block:
LOCUS AC073252 176092 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-115P16, WORKING DRAFT
SEQUENCE, 33 unordered pieces.
ACCESSION AC073252
VERSION AC073252.1 GI:8468959
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 176092)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176092)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

COMMENT

MO 63108, USA

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0115P16
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 156485 bases at least Q40
Consensus quality: 163650 bases at least Q30
Consensus quality: 165907 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 172892; sum-of-contigs
Quality coverage: 3.10 in Q20 bases; sum-of-contigs
Quality coverage: 3.50 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces is
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1718: contig of 1718 bp in length
* 1719 1818: gap of unknown length
* 1819 3309: contig of 1491 bp in length
* 3310 3409: gap of unknown length
* 3410 4912: contig of 1503 bp in length
* 4913 5012: gap of unknown length
* 5013 6111: contig of 1099 bp in length
* 6112 6211: gap of unknown length
* 6212 7785: contig of 1574 bp in length
* 7786 7885: gap of unknown length
* 7886 10408: contig of 2523 bp in length
* 10409 10508: gap of unknown length
* 10509 11959: contig of 1451 bp in length
* 11960 12059: gap of unknown length
* 12060 14663: contig of 2604 bp in length
* 14664 14764: gap of unknown length
* 14765 17346: contig of 2583 bp in length
* 17347 17446: gap of unknown length
* 17447 20149: contig of 2703 bp in length
* 20150 20249: gap of unknown length
* 20250 22576: contig of 2327 bp in length
* 22577 22676: gap of unknown length
* 22677 25605: contig of 2929 bp in length
* 25606 25705: gap of unknown length
* 25706 28264: contig of 2559 bp in length
* 28265 32264: gap of unknown length
* 32265 32365: contig of 3900 bp in length
* 32366 32665: gap of unknown length
* 32666 35877: contig of 3513 bp in length
* 35878 35977: gap of unknown length
* 35978 40783: contig of 4806 bp in length
* 40784 40883: gap of unknown length
* 40884 45988: contig of 5105 bp in length
* 45989 46088: gap of unknown length
* 46089 50028: contig of 3540 bp in length
* 50029 50128: gap of unknown length
* 50129 53931: contig of 3803 bp in length
* 53932 54031: gap of unknown length
* 54032 59121: contig of 5090 bp in length
* 59122 59221: gap of unknown length
* 59222 64713: contig of 5492 bp in length
* 64714 64813: gap of unknown length
* 64814 69477: contig of 4664 bp in length
```

```
* 69478 69577: gap of unknown length
* 69578 75283: contig of 5706 bp in length
* 75284 75383: gap of unknown length
* 75384 82227: contig of 6844 bp in length
* 82228 82327: gap of unknown length
* 82328 89905: contig of 7578 bp in length
* 89906 90005: gap of unknown length
* 90006 99898: contig of 9893 bp in length
* 99899 99999: gap of unknown length
* 99999 108174: contig of 8176 bp in length
* 108175 108275: gap of unknown length
* 108275 116628: contig of 8354 bp in length
* 116629 116729: gap of unknown length
* 116729 127264: contig of 10536 bp in length
* 127265 127365: gap of unknown length
* 127365 138142: contig of 10778 bp in length
* 138143 138243: gap of unknown length
* 138243 150067: contig of 11825 bp in length
* 150068 150167: gap of unknown length
* 150168 162076: contig of 11909 bp in length
* 162077 162177: gap of unknown length
* 162177 176092: contig of 13916 bp in length.
Location/Qualifiers
1. 176092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-115P16"
1. 1718
/misc_feature /note="assembly_name:Contig12"
/misc_feature /note="assembly_name:Contig13"
/misc_feature /note="assembly_name:Contig14"
/misc_feature /note="assembly_name:Contig15"
/misc_feature /note="assembly_name:Contig16"
/misc_feature /note="assembly_name:Contig17"
/misc_feature /note="assembly_name:Contig18"
/misc_feature /note="assembly_name:Contig19"
/misc_feature /note="assembly_name:Contig20"
/misc_feature /note="assembly_name:Contig21"
/misc_feature /note="assembly_name:Contig22"
/misc_feature /note="assembly_name:Contig23
clone_end:SP6
vector_side:right"
25706..28264
/misc_feature /note="assembly_name:Contig24"
28365..32264
/misc_feature /note="assembly_name:Contig25"
32365..35877
/misc_feature /note="assembly_name:Contig26"
35978..40783
/misc_feature /note="assembly_name:Contig27"
40884..45988
/misc_feature /note="assembly_name:Contig28"
46089..50028
/misc_feature /note="assembly_name:Contig29"
50129..53931
/misc_feature /note="assembly_name:Contig30"
54032..59121
/misc_feature /note="assembly_name:Contig31"
59222..64713
/misc_feature /note="assembly_name:Contig32"
64814..69477

FEATURES
source
1. 176092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-115P16"
1. 1718
/misc_feature /note="assembly_name:Contig12"
/misc_feature /note="assembly_name:Contig13"
/misc_feature /note="assembly_name:Contig14"
/misc_feature /note="assembly_name:Contig15"
/misc_feature /note="assembly_name:Contig16"
/misc_feature /note="assembly_name:Contig17"
/misc_feature /note="assembly_name:Contig18"
/misc_feature /note="assembly_name:Contig19"
/misc_feature /note="assembly_name:Contig20"
/misc_feature /note="assembly_name:Contig21"
/misc_feature /note="assembly_name:Contig22"
/misc_feature /note="assembly_name:Contig23
clone_end:SP6
vector_side:right"
25706..28264
/misc_feature /note="assembly_name:Contig24"
28365..32264
/misc_feature /note="assembly_name:Contig25"
32365..35877
/misc_feature /note="assembly_name:Contig26"
35978..40783
/misc_feature /note="assembly_name:Contig27"
40884..45988
/misc_feature /note="assembly_name:Contig28"
46089..50028
/misc_feature /note="assembly_name:Contig29"
50129..53931
/misc_feature /note="assembly_name:Contig30"
54032..59121
/misc_feature /note="assembly_name:Contig31"
59222..64713
/misc_feature /note="assembly_name:Contig32"
64814..69477

alignment_scores:
Quality: 86.50 Length: 72
Ratio: 2.544 Gaps: 5
Percent Similarity: 47.222 Percent Identity: 36.111

alignment_block:
US-09-599-087-5 x AC073252 ..
Align seg 1/1 to: AC073252 from: 1 to: 176092

14 LeuLeuCySpheSerIlePheSerThrIugLySarGArgProAlaLy 30
|||||
9172 TTGCTCTGCATGTCATTC.....CTGCGCG 9197

30 salAtrpsergly.....AARGYT 37
|||||
9198 CGCCTGGAGAGATGCTGCGAGAGGGGTGCACACAGACAGAGAGC 9247

37 hrArgLeuCySpheSerIlePheSerThrIugLySarGArgProAlaLy 50
|||||
9248 TCAGCCTTTCGT.....GCCCTACGACGACACTGCTCTTCGCG 9291

51 AsnLeuLySgLyHisIleValArgLeuCyLySProCyLySLeuGluPr 67
|||||
9292 AATGTT.....CTTGTCTCTCTGTCCTGTCGCCGACAGACC 9323

67 OGluProArgLeuTrp 72
|||||
9324 TCAGCTCCTCTGTGTG 9339

seq_name: gb_hgt1:AC004394

seq_documentation_block:
LOCUS AC004394 170000 bp DNA HTG 09-APR-1998
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 23
unordered pieces.
ACCESSION AC004394
VERSION AC004394.1 GI:3041827
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170000)
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
TITLE
```



```

Align seg 1/1 to reverse of: ACC004579 from: 1 to: 170000

      4  LeuValIeuSerSerIeuLeuCysIleIeuLeuLeuCysPheSerIlePh 20
      |||:||||| :|||: |||:| |||:| |||:| |||:| |||:| |||:|
124118  CTGACTTTCGCCACCCCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 124066
      20  eSerIeGluGluGlyAsnArgArgPro.AlaAlaIaIaTspSerGlyArg 36
      |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
124068  AGGGAGCCCCAGACCCACCCAGCCCTCGGC.....TGGGTTCGGCTCT 124022
      37  ThrArgLeuCysCys...HisArgVal..... 44
      |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
124024  TCCGGTCTCTGTGTGCAGTACACCGCTCTCCTGGGCGCTGTGTGCATTTTA 123975
      45  .....ProSerProAsnSerThrAsnLeuIysIleHis... 55
      |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
123974  CACAGAGGGGGGTTCCTCTCCCCMACCCC.....CGGGGTACTGTCT 123933
      56  ..His.ValArgLeuCysLysProCysLysLeuIleuProGluProArgLe 71

```

```

alignment_scores:
    Quality:      86.00      Length:      93
    Ratio:        1.792      Gaps:      7
    Percent Similarity: 51.613      Percent Identity: 34.409

alignment_block:
    US-09-599-087-5 x AC004580 ..

Align seg 1/1 to: AC004580 from: 1 to: 190000

4 LeuValLeuSerSerLeuLeuCysIleLeuLeuLeuLeuLeuCysPheSerIleLeu 20
|||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
37081 CGACGCTTTCGCCACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27130

20 eSerThrGluGluGlyLysArgArgPro..AlaLysAlaTyrSerGlyArgArg 36
::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
27131 AAGGACCCAGACCCACGACGAGGCTGCGC.....TGGGGTCCCGCTCT 27174

37 ThrArgLeuCysCys...HisArgVal..... 44

```

```

      :: ||||| |||||
27175 TCCGGTCTCTGCTCAGTCACCGCCTGCTGGGGCCTGTGTGCCATTTTA 27224
      45 .....ProSerProAsnSerThrAsnLeuLysGlyHis.... 55
      27225 CACAGGGGGTGGTCCCTCCCAACCC.....CGGGGTCACTGCT 27265
      56 ..His.ValArgLeuLysLysProCysLysLeuGluProGluProArgLe 71
      27266 GGTGGTGTCTCTGCTGCTGCCAAGTGCCTTGGAAGGAAAGAACCTGAGAG 27315
      71 uTrpValValProGluValLeuPro 79
      27316 ATGGCTCAGCCCAAGCCACCTTCCA 27340

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 03:56:24 ; Search time 44.73 Seconds
(Without Alignments)
103.515 Million cell updates/sec

Title: US-09-599-087-5

Perfect score: 442

Sequence: 1 MRLVLSLLCILLICFSIF.....PCKLEPPRLMTVPGALPOV 81

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	70.5	16.0	523	19	W78915	Bovine butyrophillin
2	69	15.6	875	19	W42011	Bovine CGMP-bindin
3	69	15.6	875	20	Y14990	Bovine CGMP-bindin
4	66	14.9	302	21	G44722	Zea mays protein f
5	65.5	14.8	90	20	Y11552	Human 5' EST sece
6	65.5	14.8	109	20	Y99424	Human PRO146 (UNQ
7	65.5	14.8	109	21	Y94964	Human secreted pro
8	65.5	14.8	109	21	Y87344	Human signal pepti
9	64.5	14.6	513	21	G21648	Arabidopsis thalia
10	64.5	14.6	513	21	G52571	Arabidopsis thalia
11	64.5	14.6	517	21	G21647	Arabidopsis thalia

12	64.5	14.6	517	21	G52570	Arabidopsis thalia
13	63	14.3	98	20	W86330	Kidney injury asso
14	63	14.3	114	9	P80681	E-L hybridprotein
15	62.5	14.1	146	11	SM04426	SMR1 polypeptide.
16	62.5	14.1	985	20	Y41716	Human PRO860 prote
17	62.5	14.1	985	21	B44272	Human PRO860 (UNQ4
18	62	14.0	192	21	Y91976	Murine interferon
19	62	14.0	586	16	R72471	Saccharomyces uvar
20	62	14.0	586	16	R72470	Saccharomyces carl
21	62	14.0	632	14	R41232	GAT-3 transporter.
22	61.5	13.9	102	20	Y38388	Human secreted pro
23	61.5	13.9	527	19	W78915	Bovine butyrophillin
24	61	13.8	90	21	B38003	Human secreted pro
25	61	13.8	91	21	B38016	Human secreted pro
26	61	13.8	91	21	B38017	Human secreted pro
27	61	13.8	159	19	W69170	N-terminally tagge
28	61	13.8	165	21	G54400	Human myelin oligo
29	61	13.8	792	21	G42010	Arabidopsis thalia
30	61	13.8	800	21	G42009	Arabidopsis thalia
31	60.5	13.7	203	16	R71361	Human truncated MO
32	60.5	13.7	247	16	R70182	Human myelin oligo
33	60.5	13.7	247	16	R71360	Human MOC. Homo s
34	60.5	13.7	247	18	W37543	Human myelin oligo
35	60.5	13.7	247	21	Y44236	Human myelin oligo
36	60.5	13.7	576	21	G46178	Arabidopsis thalia
37	60.5	13.7	583	21	G46177	Arabidopsis thalia
38	60.5	13.7	711	21	G46176	Arabidopsis thalia
39	60	13.6	477	11	R06458	v-PA.alpha2. Desm
40	60	13.6	477	11	R05122	Bat-PA(h). Desmod
41	60	13.6	875	16	R66564	Cyclic guanostine m
42	59.5	13.5	73	21	G00379	Human secreted pro
43	59.5	13.5	120	20	Y35562	Chlamydia pneumoni
44	59.5	13.5	498	13	R28805	Soluble HIV gp120.
45	59.5	13.5	509	14	R38210	LD203D4 EGF recept

ALIGNMENTS

RESULT 1	W78915	standard; Protein; 523 AA.
XX	W78915:	
AC	XX	
XX	XX	
DT	21-DEC-1998	(first entry)
XX	XX	
DE	Bovine butyrophillin protein BTF2.	
XX	XX	
KW	Bovine butyrophillin; BT; human hereditary haemochromatosis; HFE;	
KW	diagnosis; Iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;	
KW	BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;	
KW	type 1 sodium transport gene.	
OS	Bos sp.	
XX	XX	
PN	W09814466-A1.	
XX	XX	
PD	09-APR-1998.	
XX	XX	
PF	30-SEP-1997;	97WO-US17658.
XX	XX	
PR	07-MAY-1997;	97US-0852495.
PR	01-OCT-1996;	96US-0724394.
XX	XX	
PA	(PROG-) PROAGENTOR INC.	
XX	XX	
PI	Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ;	
XX	Tsuchihashi Z, Wolff RK;	
DR	WPI: 1998-240014/21.	
DR	N-PSDB: V57905.	
XX	XX	

PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0158293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160811.
 PR 21-OCT-1999; 99US-0160812.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 14.9%; Score 66; DB 21; Length 302;
 Best Local Similarity 37.3%; Pred. No. 6.1;
 Matches 22; Conservative 5; Mismatches 28; Indels 4; Gaps 2;

Oy 23 EGRKRRARAWSGRRRLCCHRVSPSPSTNLKGHHVRLCKPCKLEPEPLWVPGALPOV 81
 Db 86 eglqplphldaahtrrsrrrrvrsprslrlrharrl-rahlpphp---pvhsalpsv 140

RESULT 5
 Y11652
 ID Y11652 standard; Protein; 90 AA.
 XX Y11652;
 AC Y11652;
 XX 16-JUN-1999 (first entry)
 DT 16-JUN-1999 (first entry)
 XX Human 5' EST secreted protein SEQ ID NO:304.
 DE Human 5' EST secreted protein SEQ ID NO:304.
 XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO9906439-A2.
 PN 11-FEB-1999.
 PD 31-JUL-1998; 98WO-IB01233.
 PF 01-AUG-1997; 97US-0904468.
 PR (GSET) GENSET.
 PA Ductect A, Dumas Milne Edwards J, Lacroix B;
 PI WPI; 1999-153700/13.
 DR N-PSDB; X40370.
 XX New nucleic acids encoding human secreted proteins - obtained from
 FT cDNA libraries derived from liver, lung, large intestine, colon,
 PT thyroid and pancreas tissue
 XX Claim 27; Page 380-381; 398pp; English.

XX X40251 to X40397 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y11533 to Y11679,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, haematopoiesis regulating activity, tissue growth/differentiation
 CC activity, reproductive hormone regulating activity, chemotactic/
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity
 CC or other activities. The products can be used in forensic, gene therapy
 CC and chromosome mapping procedures. The sequences can also be used for
 CC obtaining corresponding promoter sequences. The nucleic acids encoding
 CC the signal peptide can be used for directing extracellular secretion of
 CC a polypeptide or the insertion of a polypeptide into a membrane, or
 CC importing a polypeptide into a cell.

SO Sequence 90 AA;


```

PR      17-NOV-1998;    98US-0108802.
PR      17-NOV-1998;    98US-0108806.
PR      17-NOV-1998;    98US-0108807.
PR      17-NOV-1998;    98US-0108867.
PR      17-NOV-1998;    98US-0108867.
PR      17-NOV-1998;    98US-0108925.
PR      18-NOV-1998;    98US-0108848.
PR      18-NOV-1998;    98US-0108849.
PR      18-NOV-1998;    98US-0108850.
PR      18-NOV-1998;    98US-0108851.
PR      18-NOV-1998;    98US-0108852.
PR      18-NOV-1998;    98US-0108858.
PR      18-NOV-1998;    98US-0108904.
XX
XX      (GETH ) GENENTECH INC.
PA
XX
XX      Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
PI
XX      WPI: 2000-237871/20.
DR
DR      N-PSDB; A37106.
XX
XX      New mammalian DNA sequences encoding transmembrane, receptor or
PT      secreted PRO polypeptides, useful for screening of potential peptide or
PT      small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX
PS      Claim 12; Fig 170; 773pp; English.
XX
CC      A37022 to A37144 encode the new isolated human transmembrane, receptor
CC      or secreted PRO polypeptides given in Y9340 to Y93462. The
CC      transmembrane and receptor PRO proteins can be used for screening of
CC      potential peptide or small molecule inhibitors of the relevant
CC      receptor/ligand interactions. The polypeptides and nucleotide sequences
CC      encoding them have various industrial applications and including uses as
CC      pharmaceutical and diagnostic agents. A37145 to A37330 represent
CC      PCR primers and hybridisation probes used in the isolation of the PRO
CC      polypeptides from the present invention.
XX
SQ      Sequence      109 AA;

Query Match          14.8%; Score 65.5; DB 21; Length 109;
Best Local Similarity 27.1%; Pred. No. 2.3;
Matches   23; Conservative   7; Mismatches   26; Indels   29; Gaps   4

OY      3 LLVSSLLCILLICSTISTEGRKRAKW-----SGRRTRLCCHAVSPSPNTNKG 54
        | | | | | : | : | | | | | | | | | | |
DB      10 Iplkvsfcisvtslylptn-----dlslwlpkpdlnsgtrtewsthtvpkpgf---- 61
        | | | | | | | | | | | | | | | | | | |

OY      55 HHVRICKPKCKLEPRLMWVGALP 79
        ||| | | | | | | | | | | | | | |
DB      62 -----aspc-----wplagavp 73
        | | | | | | | | | | | | | | | | | | |

RESULT       7
Y94964
ID ID      Y94964 standard; Protein; 109 AA.
XX
XX      Y94964;
AC
XX
XX      16-JUN-2000 (first entry)
DT
XX
XX      Human secreted protein clone qy442_2 protein sequence SEQ ID NO:134.
DE
XX
XX      Human; secreted protein; immunostimulant; immunosuppressant; virucide;
KW      antibacterial; antifungal; cytosstatic; antiinflammatory; dermatological;
KW      antidiabetic; antiallergic; antiarthritic; antirheumatic; protozoicide;
KW      antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW      infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW      connective tissue disease; multiple sclerosis; erythematosis;
KW      rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW      Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW      insulin dependent diabetes mellitus; graft-versus-host-disease;
KW      autoimmune inflammatory eye disease; allergy.
XX
```

```

OS Homo sapiens.
XX WO200009552-A1.
PN
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GENY ) GENETICS INST INC.
PI Jacobs K, McCoy JM, Lavalie ER, Collins-Racie LA, Evans C;
PI Metberg D, Treacy M, Agostino MC, Steinsinger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
PS Claim 143; Page 595; 641pp: English.
XX
CC A1618 to A1697 encode the human secreted proteins given in Y94898
CC to Y94980, isolated from human adult brain, adult thymoid, adult retina,
CC foetal carcinoma, adult blood, adult neural, foetal kidney, adult
CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal
CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and
CC adult bladder. cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such asthma. A16698 to A16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 109 AA:
Query Match 14.8%; Score 65.5; DB 21; Length 109;
Best Local Similarity 27.1%; Pred. No. 2.3;
Matches 23; Conservative 7; Mismatches 26; Indels 29; Gaps 4;
OY 3 LIVSSLLCILLICSTSTEGKRRAKAW-----SGRRTRLCCHRVSPNSTNLKG 54
Db 10 lptkvsfcvstslvte---dlajwlpkpdlnsgtstvshlvysktpat---- 61
OY 55 HHVRICKPKCKLEPERLWVPGALP 79
Db 62 -----aspc-----wplagavp 73

```


PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0138452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141843.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148655.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151368.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155485.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0158765.
PR 07-OCT-1999; 99US-0158829.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160811.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 21-JUN-1999; 99US-0139763.
 PR 22-JUN-1999; 99US-0139817.
 PR 23-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 28-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 01-JUL-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142055.
 PR 08-JUL-1999; 99US-0142390.
 PR 09-JUL-1999; 99US-0142803.
 PR 12-JUL-1999; 99US-0142920.
 PR 13-JUL-1999; 99US-0142977.
 PR 14-JUL-1999; 99US-0143542.
 PR 15-JUL-1999; 99US-0143624.
 PR 16-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 19-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 26-JUL-1999; 99US-0145224.
 PR 27-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148686.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159295.
 PR 13-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161921.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 14.6%; Score 64.5; DB 21; Length 517;
 Best Local Similarity 39.6%; Pred. No. 16;
 Matches 19; Conservative 6; Mismatches 14; Indels 9; Gaps 3;
 Oy 9 LICIT-LICFSTFTEGRRPAKMSGRRTLCRHVPSPNSTNKGH 55
 Db 18 lllcflsvcslyff----tkp-----srrdangcdlpspslpvigh 57
 RESULT 12
 GS2570
 ID GS2570 standard; Protein; 517 AA.
 XX

AC G52570;
XX
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 66843.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130444.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136783.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139110.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144088.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144881.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150366.

PR	26-AUG-1999;	99US-015084.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-015106.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154039.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158366.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

[illegible]

KW KIM; tissue growth promotion; regeneration; renal condition;
 KW acute renal failure; acute nephritis; tumour.
 XX
 OS Rattus sp.
 XX
 PN WO9853071-A1.
 XX
 PD 26-NOV-1998.
 XX
 PE 22-MAY-1998; 98WO-US10547.
 XX
 PR 23-MAY-1997; 97US-0047491.
 PR 23-MAY-1997; 97US-0047490.
 XX
 PA (BIOJ) BIOGEN INC.
 XX
 PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
 XX
 DR WI; 1999-045312/04.
 DR N-PSDB; V80618.
 XX
 PT kidney injury-associated molecule, KIM, polypeptides - upregulated
 PT in injured or regenerating tissues, useful to promote tissue growth
 XX and regeneration, especially to treat renal conditions
 XX
 PS Claim 17; Page 153; 213pp; English.

The present sequence represents a kidney injury associated molecule (KIM) protein. KIM proteins can be administered therapeutically by expressing KIM encoding polynucleotides, to promote growth and/or survival of damaged tissue (e.g. renal tissue), since the KIM proteins are upregulated in injured or regenerating (especially renal) tissues. KIM fusion proteins, conjugates, antibodies and vectors can also be used therapeutically, e.g. these or the KIM proteins may be included with an acceptable carrier in pharmaceutical compositions, useful for therapy/prophylaxis of conditions associated with dysfunction/dysregulation of KIM genes or proteins, especially renal diseases or impairments of renal function in humans (e.g. acute renal failure, acute nephritis). The polynucleotides can be used to produce antisense sequences which, when internalised into cells, can disrupt expression of a cellular KIM gene, also useful in therapy (e.g. to block the growth of tumours dependent on KIM for growth) or compositions. The proteins and polynucleotides are useful diagnostically e.g. to detect and quantify renal injury/disease (indicative of increased risk, or presence of, renal injury or impaired function), or abnormal responses to tissue injury (indicative of increased risk, or presence of, an autoimmune response or abnormal tissue growth arising from/affecting renal tissue). The proteins can also be used to locate KIM-producing cells (especially specific loci, e.g. tissue masses abnormally producing/expressing KIM such as tumours arising from/affecting renal tissue), by contacting cells with an imageable KIM-binding reagent and imaging reagent accumulation.

```

Query Match Similarity 14.3%; Score 63; DB 20; Length 98;
Best Local Similarity 24.0%; Pred. No. 4;
Matches 18; Conservative 16; Mismatches 27; Indels 14; Gaps 3;

OY 8 SLICILLICSTFSEEGK-RRPAKAWSGRRTRLCCHRVSPNSNTNLKGHHVRLCKPCKL 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24 sqclcltlvvaillpseagtpkexrsftahgsvtlh-vvpvddts-----pav 71
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 66 EEPRLMVVPGALPQ 80
   : | : | | | | | |
Db 72 qtlpplwtseakp 86
   : | : | | | | | |

RESULT 14
P80681
ID P80681 standard; protein; 114 AA.
XX
AC P80681;

```


[illegible]

XX 19-APR-1990.
PD
XX
XX 11-OCT-1989; 89MO-FR00523.
PF
XX
PR 11-OCT-1988; 88FR-0013353.
XX
PA (INSP) INST PASTEUR.
XX
PL Chuplin I, Tronik D, Rougeon F, Seidah N;
XX
DR WP1, 1990-147823/19.
DR N-PSDB; Q04272.
XX
PT New polypeptide useful therapeutically and in diagnosis - isolated from
PT rat submaxillary gland and derived tetra- or penta-peptide(s), antibodies
PT and hybridomas
XX
PS Claim 9 ; page 18; 26pp; French.
XX
CC This is the sequence of SMRI polypeptide which is secreted from rodent
CC esp. rat submaxillary glands and is associated with control of behaviour
CC in such animals. Abs derived from SMRI are useful for detecting its
CC presence in biological tissues and fluids.
CC See also R04389 and R04420-25.
XX
SQ Sequence 146 AA;

Query Match 14.1%; Score 62.5; DB 11; length 146;
Best Local Similarity 34.9%; Pred. No. 7/
Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6

OY 1 MRLLVSLILILLICFIFTEGRRRAKAMSGRTR--LCGRV---PSPNSTNLKG 54
Db 1 mkxlylllgllllaclq--sggyrtprrtgmprdqpsltphylgldpdngsqql-- 56

OY 55 HHVRICKCKLEPEPRLMV-VPG 76
Db 57 -gvctilplnlqp-prvlvnlpg 77

```
Search completed: June  3, 2001, 04:35:32
Job time: 2348 sec
```

RESULT	15
R04426	
ID	R04426 standard; protein; 146 AA
XX	
XX	R04426;
XX	
AC	
DT	21-SEP-1989 (first entry)
XX	
DE	SMR1 polypeptide.
XX	
XX	
KW	Submaxillary gland; behavioural
XX	
OS	Rattus.
XX	
XX	
PN	W09003981-A.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:16:14 ; Search time 120.83 Seconds
(without alignments)
3894.130 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806

Sequence: 1 ggaacgaggaataatctgcc.....atccatgcagacacacaaaa 806

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Genesec_0401.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	770.8	95.6	801	21	A72224
C 2	754.8	93.6	797	21	A72230
C 3	36.4	4.5	425	21	C43732
C 4	35.4	4.4	729	21	C58879
C 5	34.6	4.3	605	21	C52622
C 6	34.6	4.3	1697	21	C77781
C 7	34.4	4.3	5099	19	V38239
C 8	34.4	4.3	10708	19	V69286
C 9	34.2	4.2	667	20	X04334
C 10	33.8	4.2	465	17	T62715
C 11	33.4	4.1	801	21	C76995

C 12	33.4	4.1	2601	21	C48719	Arabidopsis thalia
C 13	33.2	4.1	2696	7	N60909	Plasmid pAU157 ins
C 14	33.2	4.1	2696	7	N60889	Plasmid pAU157 seq
C 15	33.2	4.1	2697	7	N60864	Sequence of Plasmid
C 16	33.2	4.1	86584	21	F22292	BAC containing rep
C 17	33	4.1	2128	21	F24275	Human PHEIX cDNA
C 18	33	4.1	3313	20	Z34221	Human PRO707 cDNA
C 19	33	4.1	3313	21	C78555	Human PRO707 nucle
C 20	32.8	4.1	932	21	C59288	Human secreted (UNQ3
C 21	32.8	4.1	6978	21	N90096	Sequence of Plasmid
C 22	32.6	4.0	1509	16	T06037	Human ALK-1 cDNA.
C 23	32.6	4.0	1509	18	T87878	Human activin rece
C 24	32.6	4.0	1984	20	T09841	Human hALK-1 clone
C 25	32.4	4.0	160	22	C89205	Human brain T calc
C 26	32.4	4.0	868	21	A65434	Porcine BAC-PIGF2-
C 27	32.4	4.0	6729	20	X83481	Human T-type volta
C 28	32.4	4.0	6750	20	X83481	Human T-type volta
C 29	32.4	4.0	6783	20	X83482	Human T-type volta
C 30	32.4	4.0	6804	20	X83483	Human T-type volta
C 31	32.4	4.0	50000	21	A64139	Nucleotide sequenc
C 32	32.4	4.0	66566	21	A53450	Human thioresoxin
C 33	32	4.0	882	21	C75955	Human ORFX ORF1510
C 34	32	4.0	892	21	F21911	Human breast and o
C 35	31.8	3.9	657	19	V55932	Human liver cancer
C 36	31.8	3.9	723	16	O79903	Human hepatoma der
C 37	31.8	3.9	1632	21	C77041	Human ORFX ORF2596
C 38	31.8	3.9	1735	21	A78385	Human secreted pro
C 39	31.8	3.9	2376	16	O79902	Human hepatoma der
C 40	31.8	3.9	2376	21	A40118	Human HDGFIN cDNA.
C 41	31.6	3.9	687	21	C79704	Human secreted pro
C 42	31.6	3.9	940	21	C77777	Human cancer assoc
C 43	31.6	3.9	1302	21	Z55555	Canine Interleukin
C 44	31.6	3.9	1302	21	Z55556	Human secreted pro
C 45	31.6	3.9	1452	20	Z00461	Human secreted pro

ALIGNMENTS

RESULT 1	A72224/c	
XX	A72224 standard; cDNA: 801 BP.	
XX		
AC	A72224:	
XX		
DT	06-DEC-2000 (first entry)	
XX		
DE	Human CASB611 cDNA.	
XX		
KM	Human; CASB611; colon-specific expression; expressed sequence tag;	
KW	ESR; colon cancer; tumour; autoimmune disease; diagnosis;	
KW	disease susceptibility; prophylaxis; genetic vaccine; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200043509-A2.	
XX		
PD	27-JUL-2000.	
XX		
PF	17-JAN-2000; 2000WO-EP00346.	
XX		
PR	19-JAN-1999; 99GB-0001078.	
PR	29-JAN-1999; 99GB-0002090.	
PR	01-FEB-1999; 99GB-0002163.	
PR	01-FEB-1999; 99GB-0002168.	
PR	01-FEB-1999; 99GB-0002169.	
PR	07-APR-1999; 99GB-0007901.	
XX		
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.	
XX		
PI	Vinals-Bassols C;	
XX		
DR	WPI; 2000-482912/42.	

PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query March 4.3%; Score 34.6; DB 21; Length 605;
Best Local Similarity 67.1%; Pred. No. 1.4; Mismatches 24; Indels 0; Gaps 0;
Matches 49; Conservative 0;

QY 35 cttctagtccttcacgcctgcctgtaacctcttccttccttccttcacatctccacaca 94
||||| ||||| ||| ||||| ||| ||| ||||| |||||
Db 368 CTTCTCTCTTCGTCACACTCTCTTATCATCATCATCTTCGCTTCATCTCTTCATG 309
||| ||||| |||
QY 95 gaagggaagagc 107
||| ||||| |||
Db 308 GAAAGGAAGCGCG 296

RESULT 6
C77781
ID C77781 standard; cDNA; 1697 BP.

AC C77781;

DT 08-FEB-2001 (first entry)

XX Human cancer associated gene sequence SEQ ID NO:175.

DE Human; cancer associated gene; cancer antigen; detection; cancer;
XX diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
KW antidiabetic; antisthmatic; antirheumatic; antirheumatic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.

OS Homo sapiens.

KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KX	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX	
OS	Homo sapiens.
NN	
PN	MO9856804-AL.
XX	
PD	17-DEC-1998.
XA	
PF	11-JUN-1998; 98WO-US12125.
XX	
PR	02-OCT-1997; 97US-0061060.
PR	13-JUN-1997; 97US-0049547.
PR	13-JUN-1997; 97US-0049548.
PR	13-JUN-1997; 97US-0049549.
PR	13-JUN-1997; 97US-0049550.
PR	13-JUN-1997; 97US-0049606.
PR	13-JUN-1997; 97US-0049607.
PR	13-JUN-1997; 97US-0049608.
PR	13-JUN-1997; 97US-0049609.
PR	13-JUN-1997; 97US-0049610.
PR	13-JUN-1997; 97US-0049611.
PR	13-JUN-1997; 97US-0050566.
PR	13-JUN-1997; 97US-0050901.
PR	13-JUN-1997; 97US-0052889.
PR	08-JUL-1997; 97US-0051919.
PR	18-AUG-1997; 97US-0055984.
PR	12-SEP-1997; 97US-0058665.
PR	12-SEP-1997; 97US-0058669.
PR	12-SEP-1997; 97US-0058669.
PR	12-SEP-1997; 97US-0058750.
PR	12-SEP-1997; 97US-0058971.
PR	12-SEP-1997; 97US-0058972.
PR	12-SEP-1997; 97US-0058975.
PR	02-OCT-1997; 97US-0060834.
PR	02-OCT-1997; 97US-0060841.
PR	02-OCT-1997; 97US-0060844.
PR	02-OCT-1997; 97US-0060865.
PR	02-OCT-1997; 97US-0061059.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Brewer LA, Ebdner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;
PI	Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;
PI	Yu GL.
DR	WPI; 1999-080881/07.
DR	P-PDB; W781149.
XX	
PT	New isolated human genes and the secreted polypeptides they encode -
PT	useful for diagnosis and treatment of e.g. cancers, neurological
PT	disorders, immune diseases, inflammation or blood disorders
PS	
XS	Claim 1; Page 189; 380pp; English.
XX	
CC	This sequence represents a nucleic acid molecule which encodes a secreted
CC	human protein. The gene number, and the clone it is derived from, are
CC	detailed in the descriptor line. The gene can be used to generate fusion
CC	proteins by linking to the gene to a human immunoglobulin Fc portion
CC	(e.g. X04307) for increasing the stability of the fused protein as
CC	compared to the human protein only.
CC	The invention relates to 86 novel genes and their fragments (nucleic acid
CC	sequences: X04311-X04410; amino acid sequences W78126-W78225) which
CC	are useful for preventing, treating or ameliorating medical conditions
CC	e.g. by protein or gene therapy. Also, pathological conditions can be
CC	diagnosed by determining the amount of the new polypeptides in a sample
CC	or by determining the presence of mutations in the new polynucleotides.
CC	Specific uses are described for each of the 86 polynucleotides, based on
CC	which tissues they are most highly expressed in (see X04311 for described
CC	uses).
XX	
XQ	Sequence 667 BP; 175 A; 165 C; 193 G; 126 T; 8 other:

```

Query Match          4.2% Score 34.2; DB 20; Length 667;
Best Local Similarity 54.8%; Pred. No. 1.9;
Matches      63; Conservative    2; Mismatches     50; Indels       0; Gaps        0
OY      692 gcatccagtggtgcccaaggaaatcccttcactgacctcgaatgatgtctgcygnaaga 751
           |||   ||   ||||   |||||   |   ||||| ||||| ||
DB      542 gaagaagagccccttcttgctccctacccctaccataaatagatgctgattctcc 601
           |||   ||   ||||   |||||   |   ||||| ||||| ||
OY      752 gcatccaacaacaagaatataataataataaactaacatgcagacacaaaa 806
           :   ||| ||| || | || ||| ||| ||| ||| || | ||| |||
DB      602 armmaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 656

RESULT 10
T62715/c
ID      T62715 standard; DNM; 465 BP.
XX
XX      T62715;
AC
XX
DT      01-MAY-1997 (first entry)
DE      DNA from construct peIF-5AM5.
XX
XX      Mutation; eukaryotic initiation factor 5A; eIF-5A; inhibition;
KW      Rev; Rex; HIV; HTLV-1; HTLV-2; viral replication; burden; ds.
XX
XX      Synthetic.
OS
XX      WO9625492-A1.
PN
PD      22-AUG-1996.
XX
XX      12-FEB-1996; 96WO-EP00585.
PF
PR      03-JUL-1995; 95GB-0013505.
PR      13-FEB-1995; 95GB-0002771.
XX
XX      (SANO ) SANDOZ LTD.
PA      (SANO ) SANDOZ PATENT GMBH.
PA      (SANO ) SANDOZ-ERFINDUNGEN VERM GMBH.
XX
XX      Hauber J;
PI
PI      WPI; 1996-393394/39.
DR      P-PADB; W00192.
XX
XX      Eukaryotic initiation factor 5A mutant protein inhibits Rev function
PT - used to treat diseases caused by retrovirus dependent on eIF-5A
PT for Rev function, partic. those caused by HIV, HTLV-I and/or HTLV-II
XX
XX      Claim 2: Page -: 32pp; English.
XX
XX      The sequences given in T62711-27 encode mutant versions of
CC eukaryotic initiation factor 5A (eIF-5A). The mutant eIF-5A
CC proteins inhibit Rev or Rex function, and can be used to treat
CC disease caused by HIV, HTLV-1 and/or HTLV-2. The repression of Rex
CC function in HTLV-1 and Rev function in HIV blocks viral replication,
CC thus lowering the viral burden by preventing the formation of
CC infective viral particles.
CC
XX
XX      Sequence 465 BP; 128 A; 106 C; 132 G; 99 T; 0 other;
SO

Query Match          4.2% Score 33.8; DB 17; Length 465;
Best Local Similarity 54.4%; Pred. No. 2.1;
Matches      68; Conservative    0; Mismatches     57; Indels       0; Gaps        0;
OY      141 ggctctgctgcacacagctcctcctcccccaactcaaaaacctgaagaagacatctgtga 200
           ||| ||||| || | ||| ||| ||| ||| ||| ||| ||| |||
DB      157 GCCCTGCGCTTGCCACTCTTCGAAAGATCATCTTCGACGCATCTTAACATGGCCGCCCTTTGA 98

```


PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135121.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 33.4; DB 21; Length 2601;
Best Local Similarity 65.1%; Pred. No. 6.5;
Matches 49; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 14 atctgcttcacacatgaagcttctagcttcacagcctgctctgctcttc 73
DB 834 ATCTTCCTTATCATCATGCTCTCTCTTCATCTTCATCTTCATCATCTCTC 775
OY 74 tgccttcacatcttc 88
DB 774 TTCTTCTCATCTTC 760

RESULT 13

N60909
ID N60909 standard; cDNA; 2696 BP.

AC N60909;

DT 13-AUG-1991 (first entry)

XX Plasmid PAU157 insert encoding rat-liver P-450MC cytochrome.

KW 3-methyl-choleanthrene; MC; ds.

OS Rattus sp.

PN JP61052284-A.

PD 14-MAR-1986.

PF 15-AUG-1984; 84JP-0169447.

PR 15-AUG-1984; 84JP-0169447.

PR 24-AUG-1984; 84JP-0175159.

XX (AGEN) AGENCY OF IND SCI TECH.

WPI: 1986-109962/17.

DR P-PSDB; P61030.

XX New plasmid used in coding rat-liver cytochrome P-450 gene.

PS Disclosure; Fig 1; 13pp; Japanese.

CC The gene product may be produced in commercial quantities from a
transformed expression system. The rat-liver cytochrome is readily
separated, and is induced by 3-methyl-choleanthrene.

SO Sequence 2696 BP; 742 A; 657 C; 624 G; 673 T; 0 other;

Query Match 4.1%; Score 33.2; DB 7; Length 2696;
Best Local Similarity 55.1%; Pred. No. 7.6;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 689 ccagcatccagtggtccccaaggaatcccttcctagcctcctgacatgagctgctgaa 748
DB 2559 caagtatcccaaaatataataggaataacgctacccgagctaaataataattacttgaa 2618

QY 749 agagcatccacaacaagaataataataaactcaactgacagacacaaaa 806
DB 2619 aa 2676

RESULT 14

N60889
ID N60889 standard; DNA; 2696 BP.

AC N60889;

DT 04-OCT-1991 (first entry)

XX Plasmid PAU157 sequence encoding rat liver P-450MC cytochrome.

KW Yeast; alcohol dehydrogenase promoter; ds.

OS Rattus sp.

PN JP61088878-A.

PD 07-MAY-1986.

PF 16-JUN-1984; 84JP-0122953.

PR 16-JUN-1984; 84JP-0122953.

PR 24-AUG-1984; 84JP-0175159.

XX (AGEN) AGENCY OF IND SCI TECH.

WPI: 1986-157863/25.

DR P-PSDB; P61082

XX Disclosure; Fig 1; 12pp; Japanese.

CC The cytochrome may be expressed by plasmid PAU157 under the control
of an upstream alcohol dehydrogenase promoter and downstream adn
terminator. The gene product may be produced from a transformed yeast

CC expression system in large quantities, and may be used for
oxidisation (fixing with NADPH-cytochrome P-450 reducing enzyme)
in a bioreactor or in waste treatment.

SO Sequence 2696 BP; 741 A; 658 C; 624 G; 673 T; 0 other;

Query Match 4.1%; Score 33.2; DB 7; Length 2696;
Best Local Similarity 55.1%; Pred. No. 7.6;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 689 ccagcatccagtggtccccaaggaatcccttcctagcctcctgacatgagctgctgaa 748
DB 2559 caagtatcccaaaatataataggaataacgctacccgagctaaataataattacttgaa 2618

QY 749 agagcatccacaacaagaataataataaactcaactgacagacacaaaa 806
DB 2619 aa 2676

RESULT 15

N60864
ID N60864 standard; DNA; 2697 BP.

AC N60864;

DT 08-JUL-1991 (first entry)

XX Sequence of plasmid PAU157 encoding rat liver cytochrome P-450MC.

KW Organic oxide removal; tac-promoter.

OS Rattus rattus.

PN JP61005783-A.

PD 11-JAN-1986.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:15:19 ; Search time 1103.61 Seconds

(without alignments)
10770.811 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806
Sequence: 1 ggaacgagggagaaatctggc.....actcaatgcagacacaaaaa 806

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 segs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pi1:*
- 13: gb_pi2:*
- 14: gb_pi3:*
- 15: gb_pi4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htgo_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_inv3:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pi:*
- 48: em_rod:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_vl:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_vil:*
- 59: gb_vil2:*
- 60: gb_vil3:*
- 61: gb_vil4:*
- 62: gb_vil5:*
- 63: gb_vil6:*
- 64: gb_vil7:*
- 65: gb_vil8:*
- 66: gb_vil9:*
- 67: gb_vil10:*
- 68: gb_vil11:*
- 69: gb_vil12:*
- 70: gb_vil13:*
- 71: gb_vil14:*
- 72: gb_vil15:*
- 73: gb_vil16:*
- 74: gb_vil17:*
- 75: gb_vil18:*
- 76: gb_vil19:*
- 77: gb_vil20:*
- 78: gb_vil21:*
- 79: gb_vil22:*
- 80: gb_vil23:*
- 81: gb_vil24:*
- 82: gb_vil25:*
- 83: gb_vil26:*
- 84: gb_vil27:*
- 85: gb_vil28:*
- 86: gb_vil29:*
- 87: gb_vil30:*
- 88: gb_vil31:*
- 89: gb_vil32:*
- 90: gb_vil33:*
- 91: gb_vil34:*
- 92: gb_vil35:*
- 93: gb_vil36:*
- 94: gb_vil37:*
- 95: gb_vil38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	770.8	95.6	801 9	AX027767 Sequence
C 2	754.8	93.6	797 9	AX027773 Sequence
C 3	482.4	59.9	2063 89	AK025416 Homo sapi
4	118	14.6	744 94	AF152002 Rattus no
5	96	11.9	742 95	S74257 2c9 gene {c
6	91.2	11.3	258746 67	AC022389 Homo sapi
7	40	5.0	1999 92	HS242859 Homo sapi
C 8	39.2	4.9	164169 72	AC041005 Homo sapi
C 9	39.2	4.9	184351 77	AC079783 Homo sapi
C 10	38	4.7	177257 73	AC068136 Homo sapi
C 11	37.8	4.7	133760 91	HS8B22 Human DNA

```
12 37.8 4.7 145981 81 AL512322 Homo sapi
13 37.8 4.7 185548 94 AC005743 Mus muscu
14 37.6 4.7 105960 91 HS209A6 AC035401 Human DNA
15 37.4 4.6 164289 76 AC079395 AC079395 Homo sapi
16 37.4 4.6 179998 78 AC087872 AC087872 Mus muscu
17 37.2 4.6 860 53 CNO50545 AL412802 t7 end of
18 37.2 4.6 107560 69 AC024567 AC024567 Homo sapi
19 37.2 4.6 128697 61 AC010236 AC010236 Homo sapi
20 37 4.6 156371 79 AL354796 AL354796 Homo sapi
21 36.8 4.6 201833 83 CNO507CL AL35102 Homo sapi
22 36.8 4.6 804 6 MOTRYPC L16807 Manduca sex
23 36.8 4.5 195964 65 AC018879 AC018879 Homo sapi
24 36.6 4.5 66771 76 AC079220 AC079220 Homo sapi
25 36.6 4.5 161868 86 AC007241 AC007241 Homo sapi
26 36.6 4.5 161159 73 AC068171 AC068171 Homo sapi
27 36.6 4.5 175140 73 AC064823 AC064823 Homo sapi
28 36.6 4.5 176894 65 AC018699 AC018699 Homo sapi
29 36.6 4.5 178780 87 AC015978 AC015978 Homo sapi
30 36.6 4.5 180485 71 AC027145 AC027145 Homo sapi
31 36.6 4.5 186153 83 CNO507ECY AL445363 Homo sapi
32 36.6 4.5 186153 83 CNO507ECY AL445363 Homo sapi
33 36.6 4.5 186439 74 AC069130 AC069130 Homo sapi
34 36.4 4.5 90890 92 HSJ31F15 AL117378 Human DNA
35 36.4 4.5 187999 71 AC026928 AC026928 Homo sapi
36 36.2 4.5 154861 66 AC019069 AC019069 Homo sapi
37 36.2 4.5 166719 80 AL355878 AL355878 Homo sapi
38 36.2 4.5 169312 64 AC016094 AC016094 Homo sapi
39 36.2 4.5 175448 65 AC018651 AC018651 Homo sapi
40 36.2 4.5 189180 68 AC023979 AC023979 Homo sapi
41 36 4.5 41055 86 AC006930 AC006930 Homo sapi
42 36 4.5 157080 80 AL358856 AL358856 Homo sapi
43 36 4.5 176607 71 AC027272 AC027272 Homo sapi
44 36 4.5 216133 78 AC087153 AC087153 Mus muscu
45 36 4.5 225358 78 AC087150 AC087150 Mus muscu
```

ALIGNMENTS

```
RESULT 1
AX027767/c 801 bp DNA PAT 16-SEP-2000
LOCUS Sequence 1 from Patent W00043509.
DEFINITION AX027767
ACCESSION AX027767
VERSION AX027767.1 GI:10188619
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 801)
AUTHORS Vinals-Bassols,C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 1 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
Source 1..801
Location/Qualifiers
1..801
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 165 a 181 c 255 g 200 t
ORIGIN
```

Query Match 95.6%; Score 770.8; DB 9; Length 801;
Best Local Similarity 99.5%; Pred. No. 2,7e-223;
Matches 794; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```
Oy 1 ggaaggaaggaatcttccttcacatagagcttccttcacagcctctg 60
|||||
Db 797 ggaaggaaggaatcttccttcacatagagcttccttcacagcctctg 738
|||||
Oy 61 tatctgcttccttccttcacatcttcacagaaaggaaggaagcctgcaa-gg 119
|||||
```

```
Db 737 TATCTGCTCTGCTTCCTTCATCTTCTCCACAGAGGGAAGAGGCGCTCCGCAACGG 678
Oy 120 ccttgtagcaggaagaaacagctctgctccaccagctccctagcaccactaaca 179
|||||
Db 677 CCTGGTCAGGAGAGAAACAGGCTCTGCTCCACACGAGTCTTACGCCCACTCAACAA 618
Oy 180 accggaaggaatcatcttgagctctgtaaacatgaaagtggagagagccccc 239
|||||
Db 617 ACTGAAAGGACATCATGTGAGCTCTGTAAACCATGCACTTGAGCCAGAGCCGCC 558
Oy 240 ttggtgtgtctggtggtgcaatccacagagtgtagcaatcccaagagagctcagaca 299
|||||
Db 557 TTTGGGTGCTGCTGGGG-AGTCCACAGAGGTGACATCTCCAAAGCAAGATCCAGACA 499
Oy 300 ggggaagaaactatgctctgacatgagtaaccagagcctctctccttcagc 359
|||||
Db 498 GCGGAGAACCTCATGCTGCGACCTGAGTACCAAGCAGCTCTGCTCCCTTCAGC 439
Oy 360 cttaacagcagtgagctgcaatgtttgagaggttcattctgggtgctgaagagccctgga 419
|||||
Db 438 CTTCACACAGCTGAGCTGCATGTTGGAGGCTTATCTCGGCTGCAAGAACCTTGGA 379
Oy 420 aagttccagaaactcagctctgtctcaatctggtccatcacttcagagctatcatga 479
|||||
Db 378 AAGTTCAGAACTCCACGCTCTGTCTCAATTGTGCCATCACTTGACACTATCATGA 319
Oy 480 gccaaactcaccacacagagcctcagtcgacacatgtggcctctcagtgcaaacac 539
|||||
Db 318 GCCAACCTCACCCACAGAGGCTCATGTCGACCATGTCGCGCTCCAGCAAAACAC 259
Oy 540 cgagcattccacacatgacagctcagctacagctacacacacacacacacacacac 599
|||||
Db 258 CGACATTCACACATGACGCTGACAGTACAAATTCAGAACCACTCATCTGCTGACAG 199
Oy 600 tgcaggtgtgcaagcaccacacaggtgtgctgacacagactgacagctcctccatctca 659
|||||
Db 198 TGCAGGGTGGCAACCAACCAAGGTGGCTGACCAAGATGACAGAGTCTCCATCTTCA 139
Oy 660 ggtccattcagctcctggtcatttaactacagcatcagtggtcccaagaatccctt 719
|||||
Db 138 GGTCCATTCACGCTCTGCTGCAATTAACATACCATGATGCTGCTCCCAAGAAATCCCTT 79
Oy 720 ccttagccttcagatagctgctgtggaagagatcccaacaaagaatataataa 779
|||||
Db 78 CCTAGCTCTCTGACATGCTGCTGGAAGAGCATTCACAAACAGATTAATATAA 19
Oy 780 taataaactcaatgcag 797
|||||
Db 18 TAAATAAATCAATGCAG 1
```

```
RESULT 2
AX027773/c 797 bp DNA PAT 16-SEP-2000
LOCUS Sequence 7 from Patent W00043509.
DEFINITION AX027773
ACCESSION AX027773
VERSION AX027773.1 GI:10188625
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS Vinals-Bassols,C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 7 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
Source 1..797
Location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 164 a 180 c 254 g 199 t
ORIGIN
```


OY 488 caccocacaggg 499
Db 574 CAGCCTTCGAG 585

RESULT 4

AF152002 744 bp mRNA ROD 29-JUN-1999
LOCUS AF152002 Rattus norvegicus unknown mRNA sequence.
DEFINITION AF152002
ACCESSION AF152002.1 GI:5257466
VERSION
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 744)
AUTHORS Rosman,T.G. and Li,P.
TITLE 2C9-like sequence expressed in lead-resistant rat glioma cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 744)
AUTHORS Rosman,T.G. and Li,P.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1999) Environmental Medicine, NYU School of
Medicine, 57 Old Forge, Tuxedo, NY 10987, USA
FEATURES
source
1. 744
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/cell_line="Pbri"
/tissue_type="glioma"
/note="lead resistant cells derived from C6 cell line;
Doizhanskaya et al., 1998, Biological Trace Element Res.
65:31-43"

misc_feature

1. 744
/note="sequence expressed in Pbri cells; not expressed in
C6 cells; sequence resembles 2C9, found in cells
overexpressing fos; Pbri cells do not overexpress fos"
BASE COUNT 208 a 216 c 146 g 174 t
ORIGIN

Query Match 14.6%; Score 118; DB 94; Length 744;
Best Local Similarity 60.8%; Pred. No. 9.9e-25;
Matches 293; Conservative 0; Mismatches 160; Indels 29; Gaps 5;

OY 24 tcacatgagctctctgctcttcacagcctgctctatctctctctctccca 83
Db 48 TCACATAGCGACTCTCAACCTCTCGGTTGTTCTTCATGCTTCTGTCGCG 107
OY 84 tctctccacagaaggagagcgctctctgccaagcctggtcagaggaacacggc 143
Db 108 TTCTCTCTCTCGAAGGAGGAAGCGTCTCGCAG-----TTCCGAAACTCAGGC 158
OY 144 tctgctgccacgagctccctagccccaactcaacaacctgaagaagcatatgtgagc 203
Db 159 CCCCTGTCACTATCTCTCTGATGCCAACCATTACCTGGAAGAAACCAACACAGAC 218
OY 204 tcttaaacatgcaagcttgagcagagcccgcccttggtgtgtctccggtgagccttc 263
Db 219 CTTGCAACCATGCA---GAAAGCTAGATCAATTATGAGGTGCTGGGCTCTCC 275
OY 264 cacaggtgta-gcactcccaagaagactcagacagcggaagactcatgctggtcac 322
Db 276 CACAGATATAGGCGCTCCCGAAGCTGGGCTCCACGAGATGAAGCGTGAATGCCATTGA 335
OY 323 ctgaggtaccacagcagcctctgtctccctttcagccttcacagcagtgagctgcaatg 382
Db 336 TGGAGAACAACTTCTGCGCCCTTACCAACTTCATGCG-----CAGAAGCTGTGACA 387
OY 383 ttgagaggtctcatctgaggtgaggaaccttggaagttcagagaccacagctctc 442

Db 388 CCAGATGTTTCACTTACAGACG-----CTGGAAGATTAAGATTCCAGCCCTC 439
OY 443 gttcgaattgtccatcaacttcagagctatctgagcaacctcaccacagggct 502
Db 440 GTTCCAAAGGTGCAACCAACCTTACAGAGTCACTATGATCCAGGCTCAGCCCAAGTCTT 499
OY 503 ca 504
Db 500 CA 501

RESULT 5

S74257 742 bp mRNA ROD 16-MAR-1995
LOCUS S74257
DEFINITION 2C9 gene (clone 2C9 insert) [rats, rat-1 fibroblast 208F,
Fos-transformed, mRNA, 742 nt].
ACCESSION S74257
VERSION S74257.1 GI:710462
KEYWORDS
SOURCE Rattus sp. Fos-transformed rat-1 fibroblast 208F.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 742)
AUTHORS Hennigan,R.F., Hawker,K.L. and Ozanne,B.W.
TITLE Fos-transformation activates genes associated with invasion
JOURNAL Oncogene 9 (12), 3591-3600 (1994)
MEDLINE 95060817
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI g1dbseq 157814] from the original journal article.
This sequence comes from Fig. 5.

FEATURES

source
1. 742
/organism="Rattus sp."
/db_xref="taxon:10118"
/gene="2C9 gene"
gene
BASE COUNT 208 a 213 c 146 g 175 t
ORIGIN

Query Match 11.9%; Score 96; DB 95; Length 742;
Best Local Similarity 59.3%; Pred. No. 4.9e-18;
Matches 286; Conservative 0; Mismatches 165; Indels 31; Gaps 6;

OY 24 tcacatgagctctctgctcttcacagcctgctctatctctctctctccca 83
Db 48 TCACATAGCGACTCTCAACCTCTCGGTTGTTCTTCATGCTTCTGTCGCG 107
OY 84 tctctccacagaaggagagcgctctctgccaagcctggtcagaggaacacggc 143
Db 108 TTCTCTCTCTCGAAGGAGGAAGCGTCTCGCAGTTCCGAAACTCAGG----- 157
OY 144 tctgctgccacgagctccctagccccaactcaacaacctgaagaagcatatgtgagc 203
Db 158 CTTGCTGTCACTATCTCTCTGATGCCAACCATTACCTGGAAGAAACCAACAGAC 216
OY 204 tctgtaaacatgcaagcttgagcagagcccgcccttggtgtgtctccggtgagccttc 263
Db 217 CTTGCAACCATGCA---GAAAGCTAGATCAATTATGAGGTGCTGGGCTCTCC 274
OY 264 cacaggtgta-gcactcccaagaagactcagacagcggaagactcatgctggtcac 322
Db 275 CACAGATATAGGCGCTCCCGAAGCTGGGCTCCACCGAGATGAAGCGTGAATGCCATTGA 334
OY 323 ctgaggtaccacagcagcctctgtctccctttcagccttcacagcagtgagctgcaatg 382
Db 335 TGGAGAACAACTTCTGCGCCCTTACCAACTTCATGCG-----CAGAAGCTGTGACA 386
OY 383 ttgagaggtctcatctgaggtgaggaaccttggaagttcagagaccacagctctc 442
Db 387 CCAGATGTTTCACTTACAGACG-----TGAAGATTAAGATTCAGGCCCTC 437

Y 443 gtctcaattgtccatcacttcagagctatcatgagccacactcaccacagggcct 502
Db 438 GTTCCAAAGTGTCACCACTTCAGAGTCACTATGATCCAGGCTCAGCCCAAGTCTT 497
Y 503 ca 504
Db 498 CA 499

RESULT 6
AC022389
LOCUS AC022389 258746 bp DNA HTG 10-MAY-2000
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT
SEQUENCE 56 unordered pieces.
ACCESSION AC022389 GI:7767724
VERSION AC022389.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 258746)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 258746)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 10, 2000 this sequence version replaced gi:7209016.

----- Genome Center
Center: Genome Therapeutics Corporation
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg137
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 201841 bases at least Q40
Consensus quality: 230471 bases at least Q30
Consensus quality: 236618 bases at least Q20
Insert size: 258746; sum-of-coverage
Quality coverage: 5.4x in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1122: contig of 1122 bp in length
*
* 1123 2241: gap of unknown length
* 1123 2241: contig of 1119 bp in length
*
* 2242 3534: gap of unknown length
* 2242 3534: contig of 1293 bp in length
*
* 3535 4881: gap of unknown length
* 3535 4881: contig of 1347 bp in length
*
* 4882 6044: gap of unknown length
* 4882 6044: contig of 1163 bp in length
*
* 6045 7094: gap of unknown length
* 6045 7094: contig of 1050 bp in length
*
* 7095 8494: gap of unknown length
* 7095 8494: contig of 1400 bp in length
*
* 8494: gap of unknown length

8495 9739: contig of 1245 bp in length
*
* 9740 11083: gap of unknown length
* 9740 11083: contig of 1344 bp in length
*
* 11084 12228: gap of unknown length
* 11084 12228: contig of 1145 bp in length
*
* 12229 13277: gap of unknown length
* 12229 13277: contig of 1049 bp in length
*
* 13278 14747: gap of unknown length
* 13278 14747: contig of 1470 bp in length
*
* 14748 15820: gap of unknown length
* 14748 15820: contig of 1073 bp in length
*
* 15821 16833: gap of unknown length
* 15821 16833: contig of 1013 bp in length
*
* 16834 17928: gap of unknown length
* 16834 17928: contig of 1095 bp in length
*
* 17929 19058: gap of unknown length
* 17929 19058: contig of 1130 bp in length
*
* 19059 20120: gap of unknown length
* 19059 20120: contig of 1062 bp in length
*
* 20121 21194: gap of unknown length
* 20121 21194: contig of 1074 bp in length
*
* 21195 22241: gap of unknown length
* 21195 22241: contig of 1047 bp in length
*
* 22242 23690: gap of unknown length
* 22242 23690: contig of 1449 bp in length
*
* 23691 24996: gap of unknown length
* 23691 24996: contig of 1306 bp in length
*
* 24997 26212: gap of unknown length
* 24997 26212: contig of 1216 bp in length
*
* 26213 27593: gap of unknown length
* 26213 27593: contig of 1381 bp in length
*
* 27594 29270: gap of unknown length
* 27594 29270: contig of 1677 bp in length
*
* 29271 31254: gap of unknown length
* 29271 31254: contig of 1984 bp in length
*
* 31255 33134: gap of unknown length
* 31255 33134: contig of 1880 bp in length
*
* 33135 34676: gap of unknown length
* 33135 34676: contig of 1542 bp in length
*
* 34677 36624: gap of unknown length
* 34677 36624: contig of 1948 bp in length
*
* 36625 38260: gap of unknown length
* 36625 38260: contig of 1636 bp in length
*
* 38261 39573: gap of unknown length
* 38261 39573: contig of 1313 bp in length
*
* 39574 42491: gap of unknown length
* 39574 42491: contig of 2918 bp in length
*
* 42492 44172: gap of unknown length
* 42492 44172: contig of 1681 bp in length
*
* 44173 45456: gap of unknown length
* 44173 45456: contig of 1284 bp in length
*
* 45457 47973: gap of unknown length
* 45457 47973: contig of 2517 bp in length
*
* 47974 49274: gap of unknown length
* 47974 49274: contig of 1301 bp in length
*
* 49275 51062: gap of unknown length
* 49275 51062: contig of 1788 bp in length
*
* 51063 53455: gap of unknown length
* 51063 53455: contig of 2393 bp in length
*
* 53456 55778: gap of unknown length
* 53456 55778: contig of 2322 bp in length
*
* 55779 57022: gap of unknown length
* 55779 57022: contig of 1244 bp in length
*
* 57023 59210: gap of unknown length
* 57023 59210: contig of 2188 bp in length
*
* 59211 60883: gap of unknown length
* 59211 60883: contig of 1673 bp in length
*
* 60884 62917: gap of unknown length
* 60884 62917: contig of 2034 bp in length
*
* 62918 65997: gap of unknown length
* 62918 65997: contig of 3080 bp in length
*
* 65998 68398: gap of unknown length
* 65998 68398: contig of 2401 bp in length

FEATURES
Source
Location/Qualifiers
1. 258746
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="CHROM 10"
/clone="RP11-124L5"
/clone_11b="RRC1-11"
BASE COUNT 73754 a 38179 c 53947 g 70833 t 33 others
ORIGIN

Query Match
Best Local Similarity 11.3%; Score 91.2; DB 67; Length 258746;
Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 gggaaatctgctctacacatgagctctagctcttcacagctctgctgatactg 67
Db 108236 GTGAAATCGCTCTTCACCATGAGCGCTTCAGCTTCACAGCTCTGTA 108295
|||
QY 68 ctctctctctctcatctcttcacagaggaag 103
Db 108296 GTTCTCTCTCTCTCATCTCTCCACAGAGGTAGG 108331
|||

RESULT 7
HSA242859 1999 bp mRNA PRI 02-DEC-1999
LOCUS HSA242859 Homo sapiens mRNA for langerin protein.
DEFINITION AJ242859
ACCESSION AJ242859.1 GI:5523278
VERSION AJ242859.1
KEYWORDS langerin protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1999)
AUTHORS Kiehlmeier,M., Ravel,O., Desutter-Dambuyant,C., Moore,K.,
Lavouat,J., Caux,C., Lebecque,S., Vincent,C., Schmitt,D.,
Langerin, a new transmembrane C-type lectin specific to Langerhans
cells, induces the formation of Birbeck granules
TITLE Unpublished
JOURNAL 2 (bases 1 to 1999)
REFERENCE Valladeau,J.M.
AUTHORS Direct Submission
JOURNAL Submitted (02-JUN-1999) Valladeau J.M., Laboratory for
Immunological Research, Schering-Plough, 27 Chemin des Peupliers,
69571 Dardilly Cedex, FRANCE
REMARK Revised by author 06-SEP-1999

FEATURES
Source
Location/Qualifiers
1. 1999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/cell_type="Langerhans cells"
/map="2p13"
48.1034
/codon_start=1
/product="langerin"
/protein_id="CAB62403.1"
/db_xref="GI:5523279"
/translacion="MTVEKEAPDHFVYDKONISLMPREPPKSGPSLVEPKPTVRA
ALICLTVALVASVLLQAVLPYRFMGITSDVTNVLKGRVNTSLDSEIKNSDGM
EAGVQIOMVNESIGYVRSOPLKLTSEYKANAIOILTRSEVSTLNAOIPELKSD
LEKASALNTKIRALQSLNENSKLKRNDILOVYSGMKYKGFYFSLPTWYS
AEQCVARNSHLTSTVSESDIEFLYKTAGCGLIYVIGLTKAGMEDMSVDDTPNNKVQ
SARFWIGEPERNAGNNEHCGRKAPSLQANNDAPCDITFLFICRPPVSEP"
1844.1849
BASE COUNT 586 a 494 c 450 g 469 t
ORIGIN

polyA_signal
BASE COUNT 586 a 494 c 450 g 469 t
ORIGIN

Query Match
Best Local Similarity 5.0%; Score 40; DB 92; Length 1999;
Matches 100; Conservative 0; Mismatches 80; Indels 2; Gaps 1;

QY 627 ctgaccagactgcagagctctctccatctcagctcagctcagctcagctcattac 686
Db 1735 CTCACCTCTCCGGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1794
|||

QY 687 taccagcatccagtggtcc--caaggaatccctctcagctcctcagatgagctgct 744
Db 1795 GGAAGCAGCCTTCAGTGCCTTCATGAATTCACCTTCAGCTCCTCAGAAATATGCT 1854
|||

QY 745 ggaagagcatcccaacaaacagtaataataataataatccatgcagacacaa 804
Db 1855 GCCTGGGTCAAGAAAAA 1914
|||

QY 805 aa 806
Db 1915 AA 1916
|||

RESULT 8
AC041005/c 164169 bp DNA HTG 22-MAY-2000
LOCUS AC041005 Homo sapiens chromosome 8 clone RP11-69M11 map 8, WORKING DRAFT
DEFINITION AC041005
ACCESSION AC041005.2 GI:8015554
VERSION AC041005.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 164169)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-69M11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 164169)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhaert,B., Brown,A., Burgett,G.,
Campopiano,A., Castle,A., Choquet,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,M., Gage,D.,
Galagan,J., Gardina,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hago,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Kartas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,J., Lehoczy,J.,
Lavin,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,

	*	93418	93517:	gap of unknown length
	*	93518	103025:	contig of 9508 bp in length
	*	103026	103125:	gap of unknown length
	*	103126	111980:	contig of 8855 bp in length
	*	111981	112080:	gap of unknown length
	*	112081	121952:	contig of 9872 bp in length
	*	121953	122052:	gap of unknown length
	*	122053	132256:	contig of 10204 bp in length
	*	132257	132356:	gap of unknown length
	*	132357	143820:	contig of 11464 bp in length
	*	143821	152809:	contig of 8889 bp in length
	*	152810	152909:	gap of unknown length
	*	152910	167529:	contig of 14620 bp in length
	*	167530	167629:	gap of unknown length
	*	167630	184351:	contig of 16722 bp in length
FEATURES	*		Location/Qualifiers	
source			1..184351	
			/organism="Homo sapiens"	
			/db_xref="taxon:3606"	
			/chromosome="8"	
			/clone="RP11-395114"	
			1..1489	
misc_feature			/note="assembly_name:Contig10"	
			1590..3500	
misc_feature			/note="assembly_name:Contig11"	
			3601..5447	
misc_feature			/note="assembly_name:Contig12"	
			5548..7087	
misc_feature			/note="assembly_name:Contig13"	
			7188..8560	
misc_feature			/note="assembly_name:Contig14"	
			8661..11225	
misc_feature			/note="assembly_name:Contig15"	
			11336..13841	
misc_feature			/note="assembly_name:Contig16"	
			13942..18052	
misc_feature			/note="assembly_name:Contig17"	
			18153..21062	
misc_feature			/note="assembly_name:Contig18"	
			21163..23532	
misc_feature			/note="assembly_name:Contig19"	
			23633..25379	
misc_feature			/note="assembly_name:Contig20"	
			25480..27374	
misc_feature			/note="assembly_name:Contig21"	
			27475..29929	
misc_feature			/note="assembly_name:Contig22"	
			30030..33309	
misc_feature			/note="assembly_name:Contig23"	
			33310..36735	
misc_feature			/note="assembly_name:Contig24"	
			36336..40063	
misc_feature			/note="assembly_name:Contig25"	
			40164..44472	
misc_feature			/note="assembly_name:Contig26"	
			44573..48556	
misc_feature			/note="assembly_name:Contig27"	
			48657..53859	
misc_feature			/note="assembly_name:Contig28"	
			53960..59209	
misc_feature			/note="assembly_name:Contig29"	
			59310..64824	
misc_feature			/note="assembly_name:Contig30"	
			64925..71110	
misc_feature			/note="assembly_name:Contig31"	
			71211..79081	
misc_feature			/note="assembly_name:Contig32"	
			79182..85622	
misc_feature			/note="assembly_name:Contig33"	
			85723..93417	
misc_feature			/note="assembly_name:Contig34"	
			93518..103025	

```

misc_feature      /note="assembly_name:Contig35"
103126. .111980
/note="assembly_name:Contig36"
misc_feature      /note="assembly_name:Contig37"
112081. .121952
/note="assembly_name:Contig38"
misc_feature      /note="assembly_name:Contig38"
122053. .122256
/note="assembly_name:Contig39
132357. .143820
/note="assembly_name:Contig39
clone_end:SP6
vector_side:right"
143921. .152809
/note="assembly_name:Contig40"
misc_feature      /note="assembly_name:Contig41"
152910. .167529
/note="assembly_name:Contig41"
167630. .184351
/note="assembly_name:Contig42"
misc_feature      /note="assembly_name:Contig42"
184351. .184351
/note="assembly_name:Contig42"
BASE COUNT      52360 a 38822 c 38341 g 51598 t 3230 others
ORIGIN

```

Query Match	4.9%	Score 39.2	DB 77	Length 184351
Best Local Similarity	51.1%	Pred. No. 1.3		
Matches	92	Conservative	0	Mismatches 86; Indels 0; Gaps 0
QY 294	cagacagcggagaaacctatgcctctgacacctgagtlaccacagacgctctctccct	353		
Db 85420	CAGGTAGCAGAGAACCCATCCCTCGCTGCTGCTACCTAACAAAGATACCCAGGCTTCCCT	85361		
QY 354	ttcagccttcacagcagtgagctgcgaatgtttgagggcttcactcgggctgcaagacc	413		
Db 85360	GACCCCTACCCGGGAGGCTCTTTGGCAGATTTGAGAAAGAAAATAAGGAAACCATGGAAG	85301		
QY 414	ctgggaaggtcccaacactccacgfcctctgtctcaattgtgcacaaactttcagagcga	473		
Db 85300	ATTCTTAAGCACTAACAATGTAATCATATGCTCTATGAAGGCTTCATATATGTCACACTAC	85241		

RESULT	10
AC068136/c	
LOCUS	AC068136 177257 bp DNA
DEFINITION	Homo sapiens chromosome 2 clone RP11-777J1, WORKING DRAFT SEQUENCE.
ACCESSION	AC068136
VERSION	AC068136.3 GI:9799886
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 177257)
Waterston, R. H.
The sequence of *Homo sapiens* clone
Unpublished
2 (bases 1 to 177257)
Waterston, R. H.
Direct Submission
Submitted (28-APR-2000) Genome Sequencing Center, Washington

COMMENT On Aug 13, 2000 this sequence version replaced g1:8571851.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H_NH0777J01
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primed ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads

```


variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 8B22.

The true right end of clone 394P21 is at 38727 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

8B22 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see [http://pacpac.med.buffalo.edu/ VECTOR: pcrPAC2](http://pacpac.med.buffalo.edu/VECTOR: pcrPAC2).

FEATURES

```
Source
1..133760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p35.1-36.21"
/clone="RPI-8B22"
/clone_lib="RPCI-1"
2601..2684
/note="14 copies 6 mer tctctc 69% conserved"
repeat_region
2602..2685
/note="42 copies 2 mer ct 70% conserved"
complement(<2608..>3094)
misc_feature
/note="match: GSS A0215085"
3953..4009
/note="19 copies 3 mer cct 74% conserved"
repeat_region
3953..4012
/note="10 copies 6 mer cctctt 73% conserved"
4163..4463
/note="AluX repeat: matches 12..311 of consensus"
4494..4565
/note="12 copies 6 mer tctctc 71% conserved"
4507..4569
/note="21 copies 3 mer cct 73% conserved"
5794..5893
/note="L2 repeat: matches 2633..2750 of consensus"
6323..6502
/note="MIR repeat: matches 57..258 of consensus"
6522..6698
/note="AluSg repeat: matches 137..313 of consensus"
6766..6864
/note="L2 repeat: matches 2610..2710 of consensus"
7594..7629
/note="18 copies 2 mer gt 83% conserved"
7723..7813
/note="AluY repeat: matches 1..98 of consensus"
7814..8114
/note="AluX repeat: matches 1..302 of consensus"
8115..8338
/note="AluY repeat: matches 98..289 of consensus"
8339..8533
/note="MIR repeat: matches 63..260 of consensus"
8703..9186
/note="MIR repeat: matches 1..505 of consensus"
9200..9486
/note="AluY repeat: matches 1..291 of consensus"
9569..9606
/note="L1ME2 repeat: matches 5850..5885 of consensus"
9748..10055
/note="AluX repeat: matches 1..308 of consensus"
10260..10555
/note="AluX repeat: matches 1..299 of consensus"
10946..11066
/note="MIR repeat: matches 36..165 of consensus"
11070..11277
repeat_region
/note="MIR repeat: matches 20..247 of consensus"
11831..11875
/note="L1M2 repeat: matches 5267..5309 of consensus"
12017..12327
/note="L1M2 repeat: matches 5309..5803 of consensus"
12384..12617
/note="MIR repeat: matches 29..258 of consensus"
12628..12926
/note="AluY repeat: matches 1..296 of consensus"
12936..13245
/note="AluSg repeat: matches 1..306 of consensus"
13255..13561
/note="AluSg repeat: matches 1..306 of consensus"
14093..14198
/note="L2 repeat: matches 2590..2698 of consensus"
14320..14466
/note="MIR repeat: matches 1..121 of consensus"
14467..14765
/note="AluY repeat: matches 1..299 of consensus"
14766..14827
/note="MIR repeat: matches 121..188 of consensus"
15213..15329
/note="MIR repeat: matches 32..146 of consensus"
15493..15528
/note="18 copies 2 mer tg 81% conserved"
15600..15894
/note="AluSg repeat: matches 1..300 of consensus"
16750..16928
/note="MIR repeat: matches 47..230 of consensus"
16983..17240
/note="AluSg repeat: matches 3..284 of consensus"
17377..17493
/note="MIR repeat: matches 20..139 of consensus"
17642..17947
/note="AluSg repeat: matches 1..303 of consensus"
17950..18231
/note="AluX repeat: matches 1..298 of consensus"
18264..18305
/note="MIR repeat: matches 65..105 of consensus"
19744..19953
/note="AluY repeat: matches 147..302 of consensus"
19954..20226
/note="AluSg repeat: matches 1..284 of consensus"
20227..20358
/note="AluY repeat: matches 1..147 of consensus"
20390..20646
/note="L1M5 repeat: matches 5901..6175 of consensus"
20647..20942
/note="AluX repeat: matches 1..296 of consensus"
20943..21068
/note="L1M5 repeat: matches 5780..5901 of consensus"
21852..22071
/note="MIR repeat: matches 6..217 of consensus"
22418..22548
/note="MIR repeat: matches 12..153 of consensus"
22552..22688
/note="MIR repeat: matches 1..124 of consensus"
23175..23280
/note="MIR repeat: matches 109..213 of consensus"
25526..25600
/note="L2 repeat: matches 2671..2747 of consensus"
25714..26023
/note="155 copies 2 mer gg 54% conserved"
26035..26349
/note="AluSg repeat: matches 1..306 of consensus"
26382..26456
/note="MIR repeat: matches 69..147 of consensus"
26663..26820
/note="MIR repeat: matches 32..194 of consensus"
27530..27723
/note="MIR repeat: matches 40..232 of consensus"
28368..28521
/note="MIR repeat: matches 4..175 of consensus"
```


misc_feature 55938..84252
/note="assembly-fragment:01906
fragment_chain:1"
misc_feature 84353..94669
/note="assembly-fragment:00291
fragment_chain:2"
misc_feature 94770..117664
/note="assembly-fragment:00428
fragment_chain:2"
misc_feature 11765..130910
/note="assembly-fragment:01794
fragment_chain:2"
misc_feature 131011..135097
/note="assembly-fragment:00411
fragment_chain:3"
misc_feature 135198..137289
/note="assembly-fragment:00860
fragment_chain:3"
misc_feature 137390..141404
/note="assembly-fragment:01844
fragment_chain:3"
misc_feature 141505..145981
/note="assembly-fragment:02390
fragment_chain:3"
BASE COUNT 38391 a 35486 c 35196 g 35702 t 1206 others
ORIGIN

Query Match 4.7%; Score 37.8; DB 81; Length 145981;
Best Local Similarity 52.9%; Pred. No. 3.4;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

OY 95 gaagggaagagcgtctgcgcgaagcgtgttaggcagagagaccagctctgcgcac 154
|||||
DB 71168 GGAGGAATGAGGGAGCGGAGCCGAGGCGCACGAACAGCTGCGCTCCCTC 71227
|
OY 155 cgaagtccttagcccaactcaacaactgtaagagacatcgttaggtctgtaacca 214
|
DB 71228 CCTTAACTCCTCCCAACAACAAACCTGAAGAAGAGCGCTGGGGAAGAAGTGAAGCCA 71287
|
OY 215 tgcagcttgagcagagccgcctcttggtg 247
|
DB 71288 GTCACTGTGGGCTCAGAGCACACCTCTGGCGG 71320
|

RESULT 13
AC005743 185548 bp DNA ROD 15-FEB-2000
LOCUS Mus musculus chromosome unknown clone rp21-657p21 strain
DEFINITION 129S6/SVEVTrac, complete sequence.
AC005743
ACCESSION AC005743.5 GI:4646254
VERSION HTG.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
AUTHORS Mus musculus PAC Clone 657p21
TITLE Unpublished
JOURNAL 2 (bases 1 to 185548)
REFERENCE Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
AUTHORS Direct Submission
TITLE Submitted (01-OCT-1998) Department of Chemistry and Biochemistry,
JOURNAL The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
REFERENCE Direct Submission
AUTHORS Submitted (08-DEC-1998) Department of Chemistry and Biochemistry,
JOURNAL

REFERENCE The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
AUTHORS OK 73019, USA
TITLE 4 (bases 1 to 185548)
JOURNAL Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Direct Submission
Submitted (08-APR-1999) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
REFERENCE Direct Submission
AUTHORS Submitted (15-APR-1999) Department of Chemistry and Biochemistry,
JOURNAL The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
REFERENCE Direct Submission
AUTHORS Submitted (17-APR-1999) Department of Chemistry and Biochemistry,
JOURNAL The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
7 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
REFERENCE Direct Submission
AUTHORS Submitted (22-APR-1999) Department of Chemistry and Biochemistry,
JOURNAL The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
8 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
REFERENCE Direct Submission
AUTHORS Submitted (15-FEB-2000) Department of Chemistry and Biochemistry,
JOURNAL The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Apr 22, 1999 this sequence version replaced gi:4587629.
COMMENT Location/Qualifiers
FEATURES
source 1..185548
/organism="Mus musculus"
/strain="129S6/SVEVTrac"
/db_xref="taxon:10090"
/chromosome="unknown"
/clone="rp21-657p21"
/clone_11b="RPCI mouse PAC library 21"
BASE COUNT 58272 a 40343 c 37771 g 49162 t
ORIGIN

Query Match 4.7%; Score 37.8; DB 94; Length 185548;
Best Local Similarity 57.0%; Pred. No. 3.5;
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 686 ctaccagcatccagtggtccccaaggaatccctcttagcctctgacatgagctgctg 745
|
DB 182702 CTCTCAGATGCTTACACACATCAAAAGCAGCCATACCTACATGAAGAGCTCAGATT 182761
|
OY 746 gaaagagcatccacaacaagaataaataaataaataaataaataaataaataa 805
|
DB 182762 AAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 182821
|
OY 806 a 806
DB 182822 A 182822

RESULT 14
HS209A6 105960 bp DNA PRI 29-APR-2000
LOCUS Human DNA sequence from clone RPL-209A6 on chromosome 6p22.1-24.3.
DEFINITION Contains STSs and GSSs, complete sequence.
ACCESSION AL035401
VERSION AL035401.23 GI:7378710
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

repeat_region      38019..38074
                    /note="L2 repeat: matches 2653..2705 of consensus"
repeat_region      40368..41178
                    /note="L1ME1 repeat: matches 5318..6163 of consensus"
repeat_region      41183..41970
                    /note="L1PA14 repeat: matches 4775..5553 of consensus"
repeat_region      41971..42381
                    /note="L1PA14 repeat: matches 5734..6149 of consensus"
repeat_region      42445..42665
                    /note="L1M4 repeat: matches 5061..5273 of consensus"
repeat_region      43158..43315
                    /note="MER5A repeat: matches 7..159 of consensus"
repeat_region      45347..45448
                    /note="MER5A repeat: matches 1..104 of consensus"
misc_feature       /note="complement(45414..45971)"
misc_feature       /note="match: GSS: Em:B52631"
misc_feature       /note="complement(45526..45807)"
misc_feature       /note="match: STS: Em:G54793"
misc_feature       /note="match: GSS: Em:B30942"
repeat_region      48380..48531
                    /note="76 copies 2 mer at 66% conserved"
misc_feature       /note="complement(48635..49059)"
repeat_region      48835..49217
                    /note="match: GSS: Em:AQ757741"
repeat_region      49235..49287
                    /note="THEIC repeat: matches 15..369 of consensus"
misc_feature       /note="L2 repeat: matches 2591..2642 of consensus"
repeat_region      49378..49854
                    /note="match: GSS: Em:AQ621055"
repeat_region      49389..49608
                    /note="MER30 repeat: matches 1..230 of consensus"
misc_feature       /note="match: GSS: Em:AQ117725"
repeat_region      50308..50706
                    /note="MLT2CB repeat: matches 1..462 of consensus"
repeat_region      50711..50758
                    /note="24 copies 2 mer to 87% conserved"
repeat_region      51298..51356
                    /note="MADBI repeat: matches 22..80 of consensus"
repeat_region      51773..52088
                    /note="L2 repeat: matches 1667..2013 of consensus"
repeat_region      52439..52791
                    /note="L2 repeat: matches 1..375 of consensus"
repeat_region      52792..53231
                    /note="L2 repeat: matches 1..375 of consensus"

Query Match      4.7% Score 37.6; DB 91; Length 105960;
Best Local Similarity 56.5%; Pred. No. 3.9;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy 675 ctggcatttaactaccagatccagtggtcccaaggaatccctcctagctcctcgaca 734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15275 CTGGCTTGAACCTTCGCGCAAGTGATGCTCCCTGTCGCCCAAAATGCTGGAA 15334

Oy 735 tgaagtcgtgtgaaagagatccaaacaaagaataataataataataactaaag 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15335 GGTGTGCTTTGAAATACTATATGACAACTGCAAAAAAATAACCTGATGT 15394

Oy 795 caga 798
    |||
Db 15395 CTGA 15398

RESULT 15
AC079395 164289 bp DNA HTG 12-OCT-2000
LOCUS Homo sapiens chromosome 2 clone RP11-219H23, WORKING DRAFT
DEFINITION SEQUENCE, 34 unordered pieces.
ACCESSION AC079395
VERSION AC079395.3 GI:10799480
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164289)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 164289)
Waterston,R.H.
Direct Submission
Submitted (31-AUG-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Oct 12, 2000 this sequence version replaced gi:9964991.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0219H23
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 144473 bases at least Q40
Consensus quality: 151522 bases at least Q30
Consensus quality: 154634 bases at least Q20
Insert size: 173000; agarose-fp
Insert size: 160989; sum-of-ctrls
Quality coverage: 3.04 in Q20 bases; agarose-fp
Quality coverage: 3.33 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1493: contig of 1493 bp in length
* 1494 1593: gap of unknown length
* 1594 2880: contig of 1287 bp in length
* 2881 2980: gap of unknown length
* 2981 4341: contig of 1361 bp in length
* 4342 4442: gap of unknown length
* 4443 5719: contig of 1278 bp in length
* 5720 5819: gap of unknown length
* 5820 7200: contig of 1361 bp in length
* 7201 7300: gap of unknown length
* 7301 8864: contig of 1564 bp in length
* 8865 8964: gap of unknown length
* 8965 10832: contig of 1688 bp in length
* 10833 10932: gap of unknown length
* 10933 12176: contig of 1244 bp in length
* 12177 12276: gap of unknown length
* 12277 13636: contig of 1360 bp in length
* 13637 13736: gap of unknown length
* 13737 15321: contig of 1585 bp in length
* 15322 15421: gap of unknown length
* 15422 17156: contig of 1735 bp in length
* 17157 17256: gap of unknown length
* 17257 19049: contig of 1793 bp in length
* 19050 19149: gap of unknown length
* 19150 22210: contig of 3061 bp in length
* 22211 22310: gap of unknown length
* 22311 23900: contig of 1590 bp in length
* 23901 24000: gap of unknown length
* 24001 26618: contig of 2618 bp in length
* 26619 26718: gap of unknown length
* 26719 29184: contig of 2466 bp in length

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:35:24 ; Search time 124.88 Seconds
(without alignments)
3767.839 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 1 ggaacggagggaatactgcc.....actcaatgcagacacaaaa 806

Sequence: 1 ggaacggagggaatactgcc.....actcaatgcagacacaaaa 806

Scoring table: OLIGO-MNC
Gapop 60.0 , Gapext 60.0

Searched: 678276 segs, 291890651 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq.0401.*
1: /SIDs2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDs2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDs2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDs2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDs2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDs2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDs2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDs2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDs2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDs2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDs2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDs2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDs2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDs2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDs2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDs2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDs2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDs2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDs2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDs2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDs2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDs2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	537	66.6	801	21	A72224
C 2	498	61.8	797	21	A72230
C 3	23	2.9	151	21	C20052
C 4	22	2.7	344	21	C23854
C 5	22	2.7	1844	21	F16260
C 6	20	2.5	186	21	C27822
C 7	20	2.5	441	21	C28871
C 8	20	2.5	452	17	T42809
C 9	20	2.5	493	21	C05785
C 10	20	2.5	493	21	T42545
C 11	19	2.4	81	16	T24093

12	19	2.4	102	16	T26325	Human gene signatu
C 13	19	2.4	141	16	T22310	Human gene signatu
C 14	19	2.4	167	21	C05350	Human secreted pro
C 15	19	2.4	198	21	C21088	Human secreted pro
C 16	19	2.4	208	21	C13546	Human secreted pro
C 17	19	2.4	229	21	A45277	Human secreted exp
C 18	19	2.4	230	21	A41574	Human secreted exp
C 19	19	2.4	240	21	C13564	Human secreted pro
C 20	19	2.4	240	21	C13564	Human secreted pro
C 21	19	2.4	352	21	C00643	Human secreted pro
C 22	19	2.4	353	20	V89188	EST clone C1240.
C 23	19	2.4	369	21	A45269	Human secreted exp
C 24	19	2.4	398	21	A74216	Lobliolly pine SSR
C 25	19	2.4	402	14	Q59630	Human brain expres
C 26	19	2.4	425	21	C57692	Arachidonic acid m
C 27	19	2.4	445	21	A74248	Lobliolly pine SSR
C 28	19	2.4	465	17	T13969	Elmeria gametocyte
C 29	19	2.4	465	17	X89273	Partial DNA sequen
C 30	19	2.4	466	17	T13968	Elmeria gametocyte
C 31	19	2.4	466	20	X89272	Partial DNA sequen
C 32	19	2.4	532	20	X37512	Human secreted pro
C 33	19	2.4	760	21	C51785	Arabidopsis thalia
C 34	19	2.4	2054	21	C59693	Human secreted pro
C 35	19	2.4	2147	20	Z77494	Human ovarian tumo
C 36	19	2.4	2362	13	A39685	Cytokine response
C 37	19	2.4	2362	17	T48098	Natural killer cel
C 38	19	2.4	2362	19	V42535	Human interleukin-
C 39	19	2.4	2362	19	V15227	Human interleukin-
C 40	19	2.4	2362	19	V15227	Heterodimeric huma
C 41	19	2.4	2362	19	Z28862	Human interleukin-
C 42	19	2.4	2635	20	V84521	Human secreted pro
C 43	19	2.4	2723	20	Z29906	CDNA encoding a SC
C 44	19	2.4	3309	21	F21886	Human breast and o
C 45	19	2.4	3865	18	T96833	Intron 3 of human

ALIGNMENTS

RESULT 1	A72224/c	
ID	A72224 standard; cDNA: 801 BP.	
XX		
AC	A72224;	
XX		
DT	06-DEC-2000 (first entry)	
XX		
DE	Human CASB611 cDNA.	
XX		
KW	Human: CASB611; colon-specific expression; expressed sequence tag;	
KW	EST; colon cancer; tumour; autoimmune disease; diagnosis;	
KW	disease susceptibility; prophylaxis; genetic vaccine; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200043509-A2.	
XX		
PD	27-JUL-2000.	
XX		
PF	17-JAN-2000; 2000WO-EP00346.	
XX		
PR	19-JAN-1999; 99GB-0001078.	
PR	29-JAN-1999; 99GB-0002090.	
PR	01-FEB-1999; 99GB-0002163.	
PR	01-FEB-1999; 99GB-0002168.	
PR	01-FEB-1999; 99GB-0002169.	
PR	07-APR-1999; 99GB-0007901.	
XX		
PA	(SMIK) SMITHKLINE BECHAM BIOLOGICALS.	
XX		
PI	Vials-Bassols C;	
XX		
DR	WPI; 2000-482912/42.	

XX New isolated polynucleotide useful for diagnosis and/or treatment of
PT colon cancer and autoimmune disease -
PS Claim 3; Page 34; 41pp: English.
XX
CC This sequence represents human CASB611 cDNA. This gene exhibits
CC colon-specific expression and is highly expressed in the rectum.
CC The invention relates to human CASB cDNA sequences CASB611, CASB500,
CC CASB501, CASB502, CASB505 and CASB507 (A72224-A72229, respectively), and
CC also to human CASB partial cDNA sequences (A72230-A72235) derived from
CC expressed sequence tags (ESTs). Expression of the human CASB genes
CC (with the exception of CASB611) is associated with colon tumours, and the
CC encoded proteins (sequences not given in the specification) represent
CC colon tumour-associated antigens. The cDNA sequences may be used in
CC diagnosing the presence or a susceptibility to a disease related to the
CC presence, expression or activity of CASB genes. Such diseases include
CC autoimmune diseases and especially colon cancer. The nucleic acid
CC sequences may also be used in genetic vaccines for the prophylaxis or
CC therapeutic treatment of colon cancer and autoimmune diseases.
XX
SQ Sequence 801 Bp; 165 A; 181 C; 255 G; 200 T; 0 other;

Query Match 66.6%; Score 537; DB 21; Length 801;
Best Local Similarity 100.0%; Pred. No. 4.4e-247;
Matches 537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 tcccaaggtgttagactcccaagaagactccagacagcggagaacctatctctggc 320
DB 537 TCCCAAGGTGTAGACTCCCAAGAAAGCAAGACTCCAGACGGAGAACTCATCTCGGC 478
QY 321 acctgagttaccagcagcctctgtctcccttcagccttcacagcagttgagctcaa 380
DB 477 ACTGAGGTATCCAGCAAGCCTCTGTCCTCCCTTCAGCCTTCACAGCAGTGAAGCTCAA 418
QY 381 ttgttgaggaggttattctcgggtcgaagagccttggaagtccagaatactcagttcc 440
DB 417 TGTGTGAGGGGTTTCAATCTCGGGCTGCAAGAGACCTGGGAAAGTTCCAGAACTCCACGCTCC 358
QY 441 ttgtctcaattgtgcatcaacttcagagatcatcatgagcacaactccacagagggc 500
DB 357 TTGTCCTCAATTGTGCTCCACTTTCAGAGTATCATGAGCAACCTCACCCCAAGGGC 298
QY 501 ctcsagtcgacacatgtggtgctctcagtgcaaacacacagagattccacatgacgg 560
DB 297 CTCAGTGCACCACTGATGGGCTCTCCAGTCAAAACCAACGAGCATTCACCATGACGG 238
QY 561 tcacagctacaataatcagaagacataactcgtctagatgagtgcaaggtggcaagaccacaa 620
DB 237 TCACAGCTACAAATCCAGAGACATCAATCTGTAAGTGTCAGGGTGGCAAGCACCACAA 178
QY 621 ggtgtgctgacaaagactgacagatctctcctcattcaggttcattcagcctctggca 680
DB 177 GGGTGGGCTGACCAAGAGTGCAGAGTCTCTCCTCATCTTCAGGTCAATTCAGCTCTGGCA 118
QY 681 tttaactacagcatcagtggtgtcccaaggaatcccttcctagcctctcctgacatgagtc 740
DB 117 TTTTACTACACAGCATCCAGTGTGCTCCCAAGGAATCCCTTCTAGCCTCTCATCATGAGTGC 58
QY 741 tgcctggaagagacatccaaacaagaagtaataataataataataataataatcag 797
DB 57 TGCTGGAAAGAGCATCCAAACAAAGATATTAATAATAATAACTCATGACAG 1

RESULT 2
A72230/c
ID A72230 standard; cDNA; 797 Bp.
XX
AC A72230;
XX
DT 06-DEC-2000. (first entry)
XX

DE Human CASB gene partial cDNA, SEQ ID NO:7.
XX
KW Human; CASB gene; overexpression; colon tumour-associated antigen;
KW expressed sequence tag; EST; colon cancer; tumour; autoimmune disease;
KW diagnosis; disease susceptibility; prophylaxis; genetic vaccine;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN NC0200043509-A2.
XX
PD 27-JUL-2000.
XX
PF 17-JAN-2000; 2000MO-EP00346.
XX
PR 19-JAN-1999; 99GB-0001078.
PR 29-JAN-1999; 99GB-0002090.
PR 01-FEB-1999; 99GB-0002163.
PR 01-FEB-1999; 99GB-0002168.
PR 01-FEB-1999; 99GB-0002169.
PR 07-APR-1999; 99GB-0007901.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals-Bassols C;
XX
DR WPI; 2000-482912/42.
XX
PT New isolated polynucleotide useful for diagnosis and/or treatment of
PT colon cancer and autoimmune disease -
XX
PS Disclosure; Page 35-36; 41pp: English.
XX
SQ Sequences A72230-A72235 represent human CASB gene partial cDNA
CC sequences which are derived from expressed sequence tags (ESTs). The
CC invention relates to human CASB cDNA sequences CASB611, CASB500, CASB501,
CC CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and also to
CC these human CASB partial cDNA sequences. Expression of the human
CC CASB genes (with the exception of CASB611) is associated with colon
CC tumours, and the encoded proteins (sequences not given in the
CC specification) represent colon tumour-associated antigens. The cDNA
CC sequences may be used in diagnosing the presence or a susceptibility to
CC a disease related to the presence, expression or activity of CASB genes.
CC Such diseases include autoimmune diseases and especially colon cancer.
CC The nucleic acid sequences may also be used in genetic vaccines for the
CC prophylaxis or therapeutic treatment of colon cancer and autoimmune
CC diseases.
XX
SQ Sequence 797 Bp; 164 A; 180 C; 254 G; 199 T; 0 other;

Query Match 61.8%; Score 498; DB 21; Length 797;
Best Local Similarity 100.0%; Pred. No. 1.9e-228;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 tcccaaggtgttagactcccaagaagactccagacagcggagaacctatctctggc 320
DB 533 TCCCAAGGTGTAGACTCCCAAGAAAGCAAGACTCCAGACGGGAGAACTCATCTCGGC 474
QY 321 acctgagttaccagcagcctctgtctcccttcagccttcacagcagttgagctcaa 380
DB 473 ACTGAGGTATCCAGCAAGCCTCTGTCCTCCCTTCAGCCTTCACAGCAGTGAAGCTCAA 414
QY 381 ttgttgaggaggttattctcgggtcgaagagccttggaagtccagaatactcagttcc 440
DB 413 TGTGTGAGGGGTTTCAATCTCGGGCTGCAAGAGACCTGGGAAAGTTCCAGAACTCCACGCTCC 354
QY 441 ttgtctcaattgtgcatcaacttcagagatcatcatgagcacaactccacagagggc 500
DB 353 TTGTCCTCAATTGTGCTCCACTTTCAGAGTATCATGAGCAACCTCACCCCAAGGGC 294
QY 501 ctcsagtcgacacatgtggtgctctcagtgcaaacacacagagattccacatgacgg 560
DB 297 CTCAGTGCACCACTGATGGGCTCTCCAGTCAAAACCAACGAGCATTCACCATGACGG 238

Db 293 CTCAGTCGCCACCAGCATGTGGGCTCTCCAGTGCAGAAACCCAGCATCCACATGACCG 234
QY 561 taacagctcaaatccagagaccatcctgtagtagtgaggtgcaagcaccga 620
Db 233 TCACAGCTACCAATCCAGGACCATCAATCTGCTAGAGTGCAGGGTGCAGACCCAA 174
QY 621 ggtgctgtaccagaactcagagatctcctcattcaggtccattcagcctccgga 680
Db 173 GGGTGGCTACCAAGACCTCAGAGTCTCTCATCTTCAGTCCATTCAGCCCTCGGCA 114
QY 681 tttaactaccagatccagtggtcccaagaatccctcctagcctcctgacatgac 740
Db 113 TTTAACTACCAATCCAGTGTGCTCCCAAGAAATCCCTCTAGCCTCCTGACATGAGTC 54
QY 741 tgcgtgaagaagacatcca 758
Db 53 TGCTGGAAGAGCATCCA 36

RESULT 3
C20052/C
ID C20052 standard; cDNA; 151 BP.
XX AC C20052;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 24127.
XX KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX PA Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PS (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI: 2000-500381/45.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 24127; 71pp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
SQ Sequence 151 BP; 28 A; 33 C; 28 G; 57 T; 5 other;

Query Match

2.9%; Score 23; DB 21; Length 151;

Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 770 aataataataaactca 792
Db 63 AATAATAATAATAATAACTCAA 41

RESULT 4

C23854/C
ID C23854 standard; cDNA; 344 BP.

XX AC C23854;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 27929.

XX KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.

XX PA Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PS (GEST) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI: 2000-500381/45.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 27929; 71pp + CD-ROM; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.

SQ Sequence 344 BP; 104 A; 63 C; 49 G; 122 T; 6 other;

Query Match

2.7%; Score 22; DB 21; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 766 aagtaataataataataa 787
Db 253 AAGTAATAATAATAATAATAA 232

RESULT 5

F16260/C
ID F16260 standard; cDNA; 1844 BP.

```

XX AC F16260;
XX XX
XX 13-MAR-2001 (first entry)
XX DE
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:695.
XX KW
XX Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
XX KW neuropeptide; cytosolic; cardioactive; immunomodulatory; muscular;
XX KW vulnery; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
XX KW antibacterial; gene therapy; neutral; immune; reproductive; renal;
XX KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX KW wound; infectious disease; ss.
XX OS
XX Homo sapiens.
XX PN
XX WO200055174-A1.
XX PD
XX 21-SEP-2000.
XX PF
XX 08-MAR-2000; 2000WO-US05988.
XX PR
XX 12-MAR-1999; 99US-0124270.
XX PA
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI
XX Rosen CA, Ruben SM;
XX DR
XX WPI: 2000-587513/55.
XX DR P-PSDB; B57057.
XX PT
XX Prostate cancer associated gene sequences, referred to as prostate
XX PT cancer antigens, useful for treatment, prevention, and diagnosis of
XX PT disorders such as prostate cancer.
XX PS
XX Claim 1; Page 1126-1127; 2338pp; English.
XX XX
XX F15566 to F15505 encode the human prostate cancer associated proteins,
XX CC called prostate cancer antigens, given in B56363 to B57302. The prostate
XX CC cancer antigens can have neuroprotective, cytosolic, cardioactive,
XX CC immunomodulatory, muscular, vulnery, gastrointestinal, nephrotoxic,
XX CC antiinfective, gynaecological and antibacterial activities, and can be
XX CC used in gene therapy. The prostate cancer antigen polynucleotides may be
XX CC used for detection of prostate cancer, chromosome identification, as
XX CC chromosome markers, and for numerous other diagnostic or research
XX CC purposes. The prostate cancer antigens may be used to treat disorders
XX CC such as neural, immune, muscular, reproductive, gastrointestinal,
XX CC pulmonary, cardiovascular, renal, and proliferative disorders, wounds,
XX CC and infectious diseases. F15506 to F16514 to B57303 represent sequences
XX CC used in the exemplification of the present invention.
XX SQ
XX Sequence 1844 BP; 531 A; 349 C; 386 G; 575 T; 3 other;
XX
XX Query Match 2.7%; Score 22; DB 21; Length 1844;
XX Best Local Similarity 100.0%; Pred. No. 0.48;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 768 gtaataataataataact 789
DB 371 GTAATAATAATAATAATACT 350
XX
XX RESULT 6
XX C27822/c
XX ID C27822 standard; cDNA; 186 BP.
XX AC C27822;
XX XX
XX 06-OCT-2000 (first entry)
XX DT
XX Human secreted protein 5' EST, SEQ ID NO: 31897.
XX DE

```

```

XX KW
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX KW
XX Homo sapiens.
XX OS
XX EP1033401-A2.
XX PN
XX 06-SEP-2000.
XX PD
XX 21-FEB-2000; 2000EP-0200610.
XX PF
XX 26-FEB-1999; 99US-0122487.
XX PR
XX (GEST ) GENSET.
XX PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX PI
XX WPI: 2000-500381/45.
XX DR
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX XX
XX Claim 1; SEQ ID 31897; 71pp + CD-ROM; English.
XX PS
XX The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX XX
XX Sequence 186 BP; 37 A; 44 C; 39 G; 63 T; 3 other;
XX
XX Query Match 2.5%; Score 20; DB 21; Length 186;
XX Best Local Similarity 100.0%; Pred. No. 4.6;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 770 aataataataataact 789
DB 167 AATAATAATAATAATACT 148
XX
XX RESULT 7
XX C28871/c
XX ID C28871 standard; cDNA; 441 BP.
XX AC C28871;
XX XX
XX 06-OCT-2000 (first entry)
XX DT
XX Human secreted protein 5' EST, SEQ ID NO: 32946.
XX DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX KW
XX Homo sapiens.
XX OS
XX EP1033401-A2.
XX PN
XX 06-SEP-2000.
XX PD
XX 21-FEB-2000; 2000EP-0200610.
XX PF

```

```

XX
PR 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Ducjert A, Giordano J;
XX WPI; 2000-500381/45.
XX
PR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 32946; 71pp + CD-ROM, English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 441 BP; 109 A; 98 C; 90 G; 144 T; 0 other;

Query Match 2.5%; Score 20; DB 21; Length 441;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 taataataataataaac 788
DB 305 TAATAATAATAATAATAAAC 286

RESULT 8
T42809 8
ID T42809 standard; DNA; 452 BP.
XX
AC T42809;
XX
DT 04-SEP-1997 (first entry)
XX
DE Polymorphic locus Q560 sequence.
XX
KW Primer; PCR; polymerase chain reaction; amplification; polymorphism;
KW genomic locus; gene mapping; paternity; maternity; children;
KW forensic science; fingerprinting; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH misc-feature 15..23
FT /tag= a
FT /note= "sequence contains between 3-9 undefined
FT repeat_region 27..74 nucleotides in this region"
FT /tag= b
FT /note= "this specification discloses that the repeat unit
FT TAA (bases 27-30; see below) is repeated 7-12
FT times; this sequence contains 12 repeats of the
FT TAA sequence".
FT repeat_unit 27..30
FT /tag= c
FT /note= "repeated 7-12 times; this sequence contains 12
FT repeats of this sequence"

```

```

FT repeat_unit 77..359
FT /tag= d
FT /rpt_type= Alu repeat
FT misc-feature 424..432
FT /tag= e
FT /note= "sequence contains between 3-9 undefined
FT nucleotides in this region"
XX
XX WO9634979-A2.
XX
XX 07-NOV-1996.
XX
XX 01-MAY-1996; 96WO-CA00275.
XX
XX 01-MAY-1995; 95US-0432023.
XX
XX (REHO-) CENT RECH HOPITAL SAINTE-JUSTINE.
XX Labuda D, Melancon SB, Tang JQ, Vanasse M;
XX WPI; 1996-506179/50.
XX
XX DNA amplification primer pair - used for simultaneous amplification
PT of multiple highly polymorphic genomic loci, partic. for
PT fingerprinting
XX
XX Example 1; Fig 1; 23pp; English.
XX
XX This is the sequence of the highly polymorphic locus Q560 which is
CC amplified by the primer pair R14B264/Q560max (T42806-7). The primers
CC also simultaneously amplify the loci Q900 (T42808) and Q120 (T42810).
CC The primers can be used for gene mapping, to assess paternity, maternity
CC and identity of children or in forensic science. In particular, they can
CC be used for the DNA fingerprinting identification of genetically related
CC or unrelated individuals.
XX
SQ Sequence 452 BP; 149 A; 92 C; 106 G; 85 T; 20 other;

Query Match 2.5%; Score 20; DB 17; Length 452;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 770 aataataataataataact 789
DB 365 aataataataataataact 384

RESULT 9
C05785/C
ID C05785 standard; cDNA; 493 BP.
XX
AC C05785;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 9860.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.

```

PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
DR
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1: SEQ ID 9860; 71pp + CD-ROM; English.
PS
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 493 BP; 98 A; 81 C; 89 G; 221 T; 4 other;

Query Match 2.5%; Score 20; DB 21; Length 493;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataaataaataaact 789
|||||
DB 339 AATAAATAAATAAATAACT 320

RESULT 10
ID 242545/c
XX 242545 standard; cDNA; 493 BP.
XX
AC 242545;
XX
DT 01-FEB-2000 (first entry)
XX
XX Human 5' EST isolated from a cDNA library SEQ ID NO:304.
DE
XX Human: 5' EST: expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX
XX Homo sapiens.
OS
XX
XX W09953051-A2.
XX
XX 21-OCT-1999.
PD
XX
XX 09-APR-1999; 99WO-IB00712.
PF
XX
XX 09-APR-1998; 98US-0057719.
PR
XX 28-APR-1998; 98US-0069047.
XX
XX (GEST) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX WPI: 2000-038446/03.
DR
XX P-PSDB: Y64931.
XX
XX Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX

PS Claim 1; Page 333-334; 837pp; English.
XX
XX 242265 to 243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. Y64651 to Y65438
CC represent the EST-related proteins corresponding to 242265 to 243072.
CC The 5' ESTs can be used for producing secreted human gene products.
CC They can be used to identify and isolate 5' untranslated regions (UTRs)
CC and upstream regulatory regions which control the location, development
CC stage, rate, and quantity of protein synthesis, as well as stability of
CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
CC obtain full length cDNA clones. The ESTs can also be used in forensic
CC procedures to identify individuals, or in diagnostic procedures to
CC identify individuals having genetic diseases resulting from abnormal
CC gene expression. The products may also be used in gene therapy protocols.
CC The nucleic acids encoding signal peptides can be used for directing
CC extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
CC The proteins encoded by the EST sequences may be useful in treating a
CC variety of human conditions. Secreted proteins have therapeutic value,
CC and the identification of new secreted proteins is valuable. 242249 to
CC 242264 and Y64644 to Y64650 represent sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 493 BP; 98 A; 81 C; 89 G; 221 T; 4 other;

Query Match 2.5%; Score 20; DB 21; Length 493;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataaataaataaact 789
|||||
DB 339 AATAAATAAATAAATAACT 320

RESULT 11
ID T24093
XX T24093 standard; cDNA to mRNA; 81 BP.
XX
AC T24093;
XX
XX 06-SEP-1996 (first entry)
XX
XX Human gene signature HUMGS06081.
DE
XX
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
XX Homo sapiens.
OS
XX
XX W09514772-A1.
XX
XX 01-JUN-1995.
PD
XX
XX 11-NOV-1994; 94WO-JP01916.
PF
XX
XX 12-NOV-1993; 93JP-0355504.
PR
XX
XX (MATS/) MATSUBARA K.
PA (OKUBO/) OKUBO K.
XX
XX Matsubara K, Okubo K;
PI
XX
XX WPI: 1995-206931/27.
DR
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
XX Claim 1; Page 1528; 2245pp; Japanese.
PS

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

Sequence 81 BP; 39 A; 16 C; 9 G; 17 T; 0 other;

Query Match 2.4%; Score 19; DB 16; Length 81;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 769 taataataataataa 787
 DB 56 taataataataataa 74

RESULT 12
 T26325
 ID T26325 standard; cDNA to mRNA; 102 BP.
 AC T26325;
 XX
 DT 16-OCT-1996 (first entry)
 XX
 DE Human gene signature HUMGS08565.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KM human; cloning; mapping; non-biased library; diagnosis; detection;
 KM cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9514772-A1.
 PD 01-JUN-1995.
 PF 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI; 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 PS Claim 1; Page 2057; 2245pp; Japanese.

CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

Sequence 102 BP; 50 A; 20 C; 12 G; 20 T; 0 other;

Query Match 2.4%; Score 19; DB 16; Length 102;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataataataataaac 788
 DB 61 aataataataataaac 79

RESULT 13
 T22310/C
 ID T22310 standard; cDNA to mRNA; 141 BP.
 AC T22310;
 XX
 DT 13-SEP-1996 (first entry)
 XX
 DE Human gene signature HUMGS03883.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KM human; cloning; mapping; non-biased library; diagnosis; detection;
 KM cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9514772-A1.
 PD 01-JUN-1995.
 PF 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI; 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 PS Claim 1; Page 1096; 2245pp; Japanese.

CC recognising different cell types.
 XX
 SQ Sequence 141 BP; 40 A; 28 C; 31 G; 40 T; 2 other;

Query Match 2.4%; Score 19; DB 16; Length 141;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 769 taataataataataa 787
 ||||||||||||||||
 DB 107 TAATAATAATAATAA 89

RESULT 14
 C05350/C
 ID C05350 standard; cDNA; 167 BP.
 XX
 AC C05350;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST; SEQ ID NO: 9425.
 XX
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PE 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 9425; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 SQ Sequence 167 BP; 27 A; 48 C; 32 G; 55 T; 5 other;

Query Match 2.4%; Score 19; DB 21; Length 167;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 ataataataataaac 788
 ||||||||||||||||

DB 51 AATAATAATAATAAAC 33

RESULT 15
 ID C21088 standard; cDNA; 198 BP.
 XX
 AC C21088;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST; SEQ ID NO: 25163.
 XX
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PE 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 25163; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 SQ Sequence 198 BP; 64 A; 48 C; 47 G; 37 T; 2 other;

Query Match 2.4%; Score 19; DB 21; Length 198;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 ataataataataaac 788
 ||||||||||||||||
 DB 165 ataataataataaac 183

Search completed: June 3, 2001, 04:34:35
 Job time: 3551 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:17:54 ; Search time 1103.59 Seconds

(without alignments)
10771.006 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806
Sequence: 1 ggaacgagggaaatctgcc.....actcaatgcagacacaaaa 806

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1283235 seqs, 7373929652 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_hcgo_hum:*
20: em_hcgo_inv:*
21: em_hcgo_rod:*
22: em_hcgo_hum1:*
23: em_hcgo_hum2:*
24: em_hcgo_hum3:*
25: em_hcgo_hum4:*
26: em_hcgo_hum5:*
27: em_hcgo_hum6:*
28: em_hcgo_hum7:*
29: em_hcgo_hum8:*
30: em_hcgo_inv1:*
31: em_hcgo_inv2:*
32: em_hcgo_other:*
33: em_hum1:*
34: em_hum2:*
35: em_hum3:*
36: em_hum4:*
37: em_hum5:*
38: em_hum6:*
39: em_hum7:*
40: em_hum8:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_hcgo1:*
61: gb_hcgo2:*
62: gb_hcgo3:*
63: gb_hcgo4:*
64: gb_hcgo5:*
65: gb_hcgo6:*
66: gb_hcgo7:*
67: gb_hcgo8:*
68: gb_hcgo9:*
69: gb_hcgo10:*
70: gb_hcgo11:*
71: gb_hcgo12:*
72: gb_hcgo13:*
73: gb_hcgo14:*
74: gb_hcgo15:*
75: gb_hcgo16:*
76: gb_hcgo17:*
77: gb_hcgo18:*
78: gb_hcgo19:*
79: gb_hcgo20:*
80: gb_hcgo21:*
81: gb_hcgo22:*
82: gb_hcgo23:*
83: gb_hcgo24:*
84: gb_hcgo25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	537	66.6	801	9	AX027767
c 2	498	61.8	797	9	AX027773
c 3	480	59.6	2063	89	AK025416
c 4	90	11.2	258746	67	AC022389
c 5	25	3.1	152245	68	AC022829
c 6	25	3.1	172084	76	AC078828
c 7	25	3.1	198057	75	AC073719
c 8	25	3.1	198777	80	AL357974
c 9	25	3.1	231443	75	AC073691
c 10	25	3.1	246962	75	AC073811
c 11	24	3.0	144818	89	AL139343


```
12 24 3 0 155986 63 AC013396 AC013396 Homo sapi
13 24 3 0 162486 72 AC044779 AC044779 Homo sapi
14 24 3 0 192519 94 MM0278435 AJ278435 Mus muscu
15 24 3 0 215516 69 AC024694 AC024694 Mus muscu
16 23 2 9 65278 81 AL365443_3 Continuation (4 of
17 23 2 9 76396 63 AC012995 AC012995 Drosophila
18 23 2 9 86945 60 AC002490 AC002490 Homo sapi
19 23 2 9 102818 85 AC003042 AC003042 Homo sapi
20 23 2 9 103606 85 AC005004 AC005004 Homo sapi
21 23 2 9 110000 81 AL365443_1 Continuation (2 of
22 23 2 9 110000 81 AL365443_2 Continuation (3 of
23 23 2 9 113020 83 HS39417 AL023585 Homo sapi
24 23 2 9 119198 91 HS845024 AL022101 Homo sapi
25 23 2 9 142971 87 AC020629 AC020629 Homo sapi
26 23 2 9 151846 90 AP001630 AP001630 Homo sapi
27 23 2 9 166385 80 AL358783 AL358783 Homo sapi
28 23 2 9 168821 81 AL391835 AL391835 Homo sapi
29 23 2 9 169434 90 AL359205 AL359205 Human DNA
30 23 2 9 169713 64 AC015919 AC015919 Homo sapi
31 23 2 9 171588 68 AC023198 AC023198 Homo sapi
32 23 2 9 178022 65 AC018348 AC033199 Homo sapi
33 23 2 9 180236 66 AC020600 AC018348 Homo sapi
34 23 2 9 183120 79 AL161789 AC020600 Homo sapi
35 23 2 9 187005 69 AC024309 AC024309 Homo sapi
36 23 2 9 214025 86 AC007882 AC024309 Homo sapi
37 23 2 9 218956 90 CENS0008 AL049831 Homo sapi
38 23 2 9 225216 69 AC024690 AC024690 Homo sapi
39 23 2 9 227856 60 AC007908 AC007908 Homo sapi
40 23 2 9 234542 93 HUAC02041 AC002041 Human Chr
41 23 2 9 340000 90 AP001748 AP001748 Homo sapi
42 22 2 7 1684 15 YSCDBR1 G35889 STS h14a121
43 22 2 7 1754 14 YSCDBR1 M62813 S.cerevisia
44 22 2 7 1754 14 YSCDBR1 Z28149 S.cerevisia
45 22 2 7 2411 15 YSCSDHA M86746 Saccharomyc
```

ALIGNMENTS

```
RESULT 1
AX027767/c AX027767 801 bp DNA PAT 16-SEP-2000
LOCUS Sequence 1 from Patent WO0043509.
DEFINITION AX027767
ACCESSION AX027767.1 GI:10188619
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 801)
AUTHORS Vinals-Bassols, C.
TITLE Novel compounds.
JOURNAL Patent: WO 0043509-A 1 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
source location/Qualifiers
1..801 /organism="Homo sapiens"
BASE COUNT 165 a 181 c 255 g 200 t
ORIGIN
```

Query Match 66.6%; Score 537; DB 9; Length 801;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 261 tccacaggtgtacgcatcccaagcaagatccagacagcggggaacctcatctcgtgc 320
DB 537 TCCACAGGTGTACGCTCCCAAGCAAGACTCCAGACGAGGAGAACCTCATCTCCGCG 478
QY 321 acctgaagttacacagcagctcctcgtctcccttcagccttcacagcagttgagctcaa 380
|||||
```

```
DB 477 ACCTGAGCTACCCAGACAGCCTCTCTGTCTCCCTTTACGCTTCACAGCATGAGCTGCAA 418
QY 381 tcttgaaggtctcatctcgcggctgcgaagacccctggaagattcccaactccacgtcc 440
DB 417 TGTGGAAGGGCTTATCTCGGGCTGCAAGGACCCCTGGGAAGTTCCAGAACTCCACGCTCC 358
QY 441 ttgtctcaattgtgcataactttcagaactatcatgagcacaacctcaaccacagggc 500
DB 357 TTGTCTCAATTGTGCATCACTTCAGAGTATCATATGAGCAACCTCACCCACAGGGC 298
QY 501 ctcaagtcacacacatgtgtggcctctccagtgcaaacaccagagattccaccatgacgg 560
DB 297 CTGAGTCCCAACATGTGGGCTCTCCAGTCAAAACCCAGCATTTCCACATGACCGG 238
QY 561 tcacagctacacatccagagacacacatcctgtctagatgtgcaaggtgcaagaccacaa 620
DB 237 TCACAGCTACCAATTCAGAGACATCATCTGCTAGAGTGTCAGAGGTGCAAGCACCA 178
QY 621 ggtgtgctgacacagactgcagagatctctccatcttcaggttcacatcagctcctggca 680
DB 177 GGGTGGCTGACCAAGACTGCAAGTCTCTTCATCTTCAGGTTCATTCAGCTCTGCA 118
QY 681 tttaactaccagcatcagtggtgtcccaaggaatcccttcctagccttcacatgagtc 740
DB 117 TTTACTACCAAGCATCTCAGTGTCTCCAGAGATCTCTTCTAGCTCTCTGACATGAGTC 58
QY 741 tgcctgaagagacatcccaacaaagtaataataataataactcaatgag 797
DB 57 TGGTGGAAAGGACATCCCAACAAAGTATATATATATATATATATATATATGATGAG 1
```

```
RESULT 2
AX027773/c AX027773 797 bp DNA PAT 16-SEP-2000
LOCUS Sequence 7 from Patent WO0043509.
DEFINITION AX027773
ACCESSION AX027773
VERSION AX027773.1 GI:10188625
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS Vinals-Bassols, C.
TITLE Novel compounds.
JOURNAL Patent: WO 0043509-A 7 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
source location/Qualifiers
1..797 /organism="Homo sapiens"
BASE COUNT 164 a 180 c 254 g 199 t
ORIGIN
```

Query Match 61.8%; Score 498; DB 9; Length 797;
Best Local Similarity 100.0%; Pred. No. 28-285;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 261 tccacaggtgtacgcatcccaagcaagatccagacagcggggaacctcatctcgtgc 320
DB 533 TCCACAGGTGTACGCTCCCAAGCAAGACTCCAGACGAGGAGAACCTCATCTCCGCG 474
QY 321 acctgaagttacacagcagctcctcgtctcccttcagccttcacagcagttgagctcaa 380
DB 473 ACCTGAGTACCAAGACAGCCTCTCTGTCTCCCTTTACGCTTCACAGCATGAGCTGCA 414
QY 381 tcttgaaggtctcatctcgggtgcgaagacccctgggaagttccagaaactccacgtcc 440
DB 413 TGTGGAAGGGCTTATCTCGGGCTGCAAGGACCCCTGGGAAGTTCAGAACTCCACGCTCC 354
QY 441 ttgtctcaattgtgcataactttcagaactatcatgagcacaacctcaccacagggc 500
|||||
```

Db 353 TTGTCATATTGTGCCATCACTTTCAGAGCTATCATGAGCAACCTCACCCACAGAGGC 294

QY 501 ctaagtcgacacatgtggcctctccagtcgaaacacccagacattccacatgacgg 560

Db 293 CTAGTCGACCATGAGGCGCTCTCCAGTGAACCCAGACCATTCACCATGACGG 234

QY 561 tcacagctacaatccagagacatcatctctgtagagtcgaggtgtggaacacccaa 620

Db 233 TCACAGCTACAAATCCAGACATCATCTCTGAGAGTCAGGAGGTCGACACCCAA 174

QY 621 ggggtgctgacacagacgtcagagtcctctccatcttcagtcacatcagctctgca 680

Db 173 GGGTGGCTGACCAAGACTGAGAGTCTCTCCATCTTCAGGTCATTCACGCTCTCGGCA 114

QY 681 ttaactacagatccagtcagtcgtccccaagaatccctctctgagctctgacatgagtc 740

Db 113 TTTAACTACACAGATCCAGAGTCTCCCAAGAAATCCCTTCCTGACCTCTGACATGAGTC 54

QY 741 tgctgaaagagatcca 758

Db 53 TGCTGAAAGACATCCA 36

RESULT 3
AK025416
LOCUS AK025416 2063 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens CDNA: FLJ21763 fls, clone COLF6967.
ACCESSION AK025416
VERSION AK025416.1 GI:10437924
KEYWORDS oligo cloning; fls (full insert sequence).
SOURCE Homo sapiens colon mucosa CDNA to mRNA, clone_11b:COLF clone:COLF6967.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (sites)
Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H., Ota, T., Suzuki, Y., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2063)
Sugano, S., Suzuki, Y., Ota, T., Odayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases, Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: 5'-5' and one pass sequencing: Department of Vitrology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES
source
1..2063
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COLF6967"
/clone_11b="COLF"
/tissue_type="colon mucosa"
/note="cloning vector pME18SFL3"

BASE COUNT 536 a 479 c 476 g 572 t
ORIGIN

Query Match 59.6%; Score 480; DB 89; Length 2063;
Best Local Similarity 100.0%; Pred. No. 1,2e-274;
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gaaatctgctctctcaacatgaggtcttagctcttccagcctctctgtctgt 69

Db 96 GAAATCTGCTCTCTCAACATGAGGCTTCTAGTCTTCTCCAGCCTGCTGTATCTGCT 155

QY 70 tctctgctctccatctctccacagaagaagagcgtccctgccaagcctggtcagg 129

Db 156 TCTCTGCTTCTCCATCTTCTCCACAGAGGAGAGAGGCGTCTGCAAGGCTGTGTAGG 215

QY 130 caggagaacacagagctctgctgcaacagagtccttagcccaactcaacaactgaaag 189

Db 216 CAGAGAGACACAGGCTGTGCTGCGCACAGAGTCCCTAGGCCCAATCAACAACTGAAAG 275

QY 190 acaatcatgtgagctctgttaaacatgacagcttgagcagagcccgcttgggtgt 249

Db 276 ACATCATGTGTAGGCTGTGTAAACCATGCAAGCTTGAAGCCAGACCCCGCTTGGGTGT 335

QY 250 gcttgaggacatcccaaggtgtagcatctccaaagaagactccagacagcggaacc 309

Db 336 GCTTGAGGACACTCCACAGGTGAGCATCCCAACCAACATCCAGACAGCGAGAAC 395

QY 310 tcatgctgacacatgagtaaccagaagcctctctctcccttccagcctcacagca 369

Db 396 TCATGCTGACACTGAGTACAGTACCCAGACGCTCTCTCTCCCTTCAGCCTTCACAGCA 455

QY 370 gtagctgcaatgttgagggcttcatctcggctgcaagagccctgggaaagttccaga 429

Db 456 GTGAGCTGCAATGTGTGAGGGCTTTCATCTCGGCTGCAAGAGACCTTGGAAGTTCAGA 515

QY 430 accccacgtctctgttcaatgtgcatcaacttccagagctatcatgagcaccctca 489

Db 516 ACTCCAGTCTCTGTCTCAATGTGCAATCTTCAGAGCTATCATGAGCCACACTCA 575

RESULT 4
AC022389
LOCUS AC022389 258746 bp DNA HTG 10-MAY-2000
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT
SEQUENCE, 56 unordered pieces.
ACCESSION AC022389
VERSION AC022389.3 GI:7767724
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 258746)
Smith, D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Unpublished
2 (bases 1 to 258746)
Smith, D.R.
Direct Submission
Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
On May 10, 2000 this sequence version replaced gi.7209016.

COMMENT

----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg337
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 201841 bases at least Q40
Consensus quality: 230471 bases at least Q30
Consensus quality: 236618 bases at least Q20
Insert size: 258746; sum-of-ctrls

Quality coverage: 5.4x in Q20 bases; sum-of-contigs

```
* NOTE: This is a working draft sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

1	1122:	contlg of 1122 bp in length
1123	2241:	contlg of 1119 bp in length
2242	3534:	contlg of 1293 bp in length
3535	4881:	contlg of 1347 bp in length
4882	6044:	contlg of 1163 bp in length
6045	7094:	contlg of 1050 bp in length
7095	8494:	contlg of 1400 bp in length
8495	9739:	contlg of 1245 bp in length
9740	11083:	contlg of 1344 bp in length
11084	12228:	contlg of 1145 bp in length
12229	13277:	contlg of 1049 bp in length
13278	14747:	contlg of 1470 bp in length
14748	15820:	contlg of 1073 bp in length
15821	16633:	contlg of 1013 bp in length
16834	17928:	contlg of 1095 bp in length
17929	19058:	contlg of 1130 bp in length
19059	20120:	contlg of 1062 bp in length
20121	21194:	contlg of 1074 bp in length
21195	22241:	contlg of 1047 bp in length
22242	23690:	contlg of 1449 bp in length
23691	24996:	contlg of 1306 bp in length
24997	26212:	contlg of 1216 bp in length
26213	27593:	contlg of 1361 bp in length
27594	292370:	contlg of 1677 bp in length
29271	31254:	contlg of 1984 bp in length
31255	33134:	contlg of 1860 bp in length
33135	34676:	contlg of 1542 bp in length
34677	36624:	contlg of 1948 bp in length
36625	38260:	contlg of 1656 bp in length
38261	39573:	contlg of 1313 bp in length
39574	42491:	contlg of 2918 bp in length
42492	44172:	contlg of 1661 bp in length

*	44173	44556:	gap of unknown length
*		contig of 1284 bp in length	
*	45457	47973:	gap of unknown length
*		contig of 2517 bp in length	
*	47974	49274:	gap of unknown length
*		contig of 1301 bp in length	
*	49275	51062:	gap of unknown length
*		contig of 1788 bp in length	
*	51063	53455:	gap of unknown length
*		contig of 2393 bp in length	
*	53456	55778:	gap of unknown length
*		contig of 2323 bp in length	
*	55779	57022:	gap of unknown length
*		contig of 1244 bp in length	
*	57023	59210:	gap of unknown length
*		contig of 2188 bp in length	
*	59211	60883:	gap of unknown length
*		contig of 1673 bp in length	
*	60884	62917:	gap of unknown length
*		contig of 2034 bp in length	
*	62918	65997:	gap of unknown length
*		contig of 3080 bp in length	
*	65998	68398:	gap of unknown length
*		contig of 2401 bp in length	
*	68399	74552:	gap of unknown length
*		contig of 6154 bp in length	
*	74553	78738:	gap of unknown length
*		contig of 4186 bp in length	
*	78739	83546:	gap of unknown length
*		contig of 4808 bp in length	
*	83547	90669:	gap of unknown length
*		contig of 7123 bp in length	
*	90670	100399:	gap of unknown length
*		contig of 9729 bp in length	
*	100399	108591:	gap of unknown length
*		contig of 8193 bp in length	
*	108592	121396:	gap of unknown length
*		contig of 12805 bp in length	
*	121397	131100:	gap of unknown length
*		contig of 9704 bp in length	
*	131101	148365:	gap of unknown length
*		contig of 17269 bp in length	
*	148370	168804:	gap of unknown length
*		contig of 20435 bp in length	
*	168805	210111:	gap of unknown length
*		contig of 41307 bp in length	
*	210112	258746:	gap of unknown length
*		contig of 48635 bp in length	

FEATURES

Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="CHROM 10"
/clone="RP11-124L5"
/clone_11b="RRC1-11"
BASE COUNT 73754 a 58179 c 55947 g 70833 t 33 others
ORIGIN

```

Query Match	11.2%	Score 90:	DB 67:	Length 258746:
Best Local Similarly	100.0%	Pred. No.	8.1e-42:	
Matches 90:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:

Qy	10	gaaaatcgcgcttcaacccaagagagccctgagctcttcacagcagtgctcgtatccacgct	69
Db	108238	GAAAATCGCCTTTCACACAGAGGCTTTCAGCTCTTCCAGCTGCTCTGTATCTGCT	10829
Qy	70	tctctgctctccaatctcttcacagaag	99
Db	108298	TCTCTGCTCTTCGATCTTCTTCACAGAGG	108327

RESULT 5

AC022829/c
 LOCUS AC022829 152245 bp DNA HTG 16-MAR-2000
 DEFINITION Homo sapiens clone RP11-121C1, WORKING DRAFT SEQUENCE, 20 unordered
 pieces
 AC022829
 AC022829.3 GI:7249268
 HTG: HTGS_PHASE1, HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 152245)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Unpublished
 2 (bases 1 to 152245)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
 Ferrelia,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McSheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
 Notman,C.H., O'Connor,T., O'Donnell,P., Olyvar,T.M., Peterson,K.,
 Piatre,N., Pisanu,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J.,
 Tittrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced gi:6984468.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 Center: Whitehead Institute/ MIT Center for Genome Research
 Genome Center
 Center code: WIRB
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 121_C1
 Center clone name: 16137
 Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 142478 bases at least Q40
 Consensus quality: 147185 bases at least Q30
 Consensus quality: 148983 bases at least Q20
 Insert size: 150000; agarose-fp
 Insert size: 150345; sum-of-contigs
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs
 NOTE: This is a 'working draft' sequence. It currently
 consists of 20 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 1283: contig of 1283 bp in length
 1284 1383: gap of 100 bp
 1384 1916: contig of 533 bp in length
 1917 2016: gap of 100 bp
 2017 3647: contig of 1631 bp in length

FEATURES
 source
 * 3648 3747: gap of 100 bp
 * 3748 5481: contig of 1734 bp in length
 * 5482 5581: gap of 100 bp in length
 * 5582 6718: contig of 1137 bp in length
 * 6719 6818: gap of 100 bp in length
 * 6819 10904: contig of 4086 bp in length
 * 10905 11004: gap of 100 bp in length
 * 11005 14393: contig of 3389 bp in length
 * 14394 14493: gap of 100 bp in length
 * 14494 18255: contig of 3762 bp in length
 * 18256 18355: gap of 100 bp in length
 * 18356 23193: contig of 4838 bp in length
 * 23194 23293: gap of 100 bp in length
 * 23294 28506: contig of 5213 bp in length
 * 28507 28606: gap of 100 bp in length
 * 28607 36308: contig of 7702 bp in length
 * 36309 36408: gap of 100 bp in length
 * 36409 43542: contig of 7134 bp in length
 * 43543 43642: gap of 100 bp in length
 * 43643 51047: contig of 7405 bp in length
 * 51048 51147: gap of 100 bp in length
 * 51148 61019: contig of 9872 bp in length
 * 61020 61119: gap of 100 bp in length
 * 61120 74018: contig of 12899 bp in length
 * 74019 74118: gap of 100 bp in length
 * 74119 84938: contig of 10820 bp in length
 * 84939 85038: gap of 100 bp in length
 * 85039 100034: contig of 14996 bp in length
 * 100035 100134: gap of 100 bp in length
 * 100135 112980: contig of 12846 bp in length
 * 112981 113080: gap of 100 bp in length
 * 113081 130270: contig of 17190 bp in length
 * 130271 130370: gap of 100 bp in length
 * 130371 152245: contig of 21875 bp in length.
 Location/Qualifiers
 1..152245
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-121C1"
 /clone_11b="RP11-11 Human Male BAC"
 1..1283
 /note="assembly-fragment"
 1384..1916
 /note="assembly-fragment"
 clone_end:17
 vector_side:left
 2017..3647
 /note="assembly-fragment"
 3748..5481
 /note="assembly-fragment"
 5582..6718
 /note="assembly-fragment"
 clone_end:sp6
 vector_side:right
 6819..10904
 /note="assembly-fragment"
 11005..14393
 /note="assembly-fragment"
 14494..18255
 /note="assembly-fragment"
 18356..23193
 /note="assembly-fragment"
 23294..28506
 /note="assembly-fragment"
 28607..36308
 /note="assembly-fragment"
 36409..43542
 /note="assembly-fragment"
 43643..51047
 /note="assembly-fragment"
 51148..61019
 /note="assembly-fragment"
 61120..74018
 /note="assembly-fragment"

misc_feature /note="assembly_fragment"
74119.84938
misc_feature /note="assembly_fragment"
85039.100034
misc_feature /note="assembly_fragment"
100135.112980
misc_feature /note="assembly_fragment"
113081.130270
misc_feature /note="assembly_fragment"
130371.152245
misc_feature /note="assembly_fragment"
/note="assembly_fragment"
BASE COUNT 47920 a 28046 c 27243 g 47134 t 1902 others
ORIGIN

Query Match 3.1% Score 25; DB 68; Length 152245;
Best Local Similarity 100.0%; Pred. No. 0.005; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 770 aataaataaataaactaatg 794
|||||
Db 43991 AATAAATAAATAAATACTAATG 43967

RESULT 6
AC078828 172084 bp DNA HTG 10-NOV-2000
LOCUS HOMO sapiens chromosome 3 clone RP11-779P15, *** SEQUENCING IN
DEFINITION PROGRESS ***, 48 unordered pieces.
AC078828 AC078828.6 GI:11128226
VERSION HTG; HTGS_PHASE1.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 172084)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrokh,S.L., Amarante,H.C., Are,J.R., Banks,T., Barbarella,J.,
Benton,J., Blum,M., Brown,E., Brown,N.C., Carron,T.F.,
Bowie,S., Biviera,M., Brown,E., Brown,N.C., Carron,T.F.,
Burke,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyne,M.D., Daborn,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthett,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollomey,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J.,
Kovar,C., Katovic,J., Kureshi,A., Landry,N., Ieal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieh,C., Liu,J., Liu,W.,
Louisegeed,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S.,
Oguchi,M., Okunnu,G., Oragunye,N., Oyeduro,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokkan,I., Rolfe,M.,
Rulz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Slisdon,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Teng,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 172084)
Worley,K.C.
Direct Submission
Submitted (04-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 9, 2000 this sequence version replaced gi:11034906.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBZM
Center clone name: RP11-779P15
Sequencing vector: M13: L08821
----- Summary Statistics
Chemistry: Dye-primer Body: 6% of reads
Chemistry: Dye-terminator Big Dye: 94% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124532 bases at least Q40
Consensus quality: 141470 bases at least Q30
Consensus quality: 150946 bases at least Q20
Estimated insert size: 149609; sum-of-ctrls estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 2.6x in Q20 bases; sum-of-ctrls estimation

* NOTE: Estimated insert size may differ from sequence length
* (see <http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.html>).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 12142: contig of 12142 bp in length
* 12143 12242: gap of unknown length
* 12243 12878: contig of 7636 bp in length
* 12878 19978: gap of unknown length
* 19978 28365: contig of 8387 bp in length
* 28365 28466: gap of unknown length
* 28466 36674: contig of 8209 bp in length
* 36674 36775: gap of unknown length
* 36775 44084: contig of 7310 bp in length
* 44084 44184: gap of unknown length
* 44184 49149: contig of 4965 bp in length
* 49149 49249: gap of unknown length
* 49249 56285: contig of 7036 bp in length
* 56285 56385: gap of unknown length
* 56385 60904: contig of 4519 bp in length
* 60904 61004: gap of unknown length
* 61004 61005: gap of unknown length
* 61005 61006: contig of 6723 bp in length
* 61006 67728: gap of unknown length
* 67728 73285: contig of 5458 bp in length
* 73285 73386: gap of unknown length
* 73386 78724: contig of 5339 bp in length
* 78724 78825: gap of unknown length
* 78825 83977: contig of 5153 bp in length
* 83977 83978: gap of unknown length
* 83978 84077: gap of unknown length
* 84077 89464: contig of 5387 bp in length
* 89464 89564: gap of unknown length
* 89564 93274: contig of 3710 bp in length
* 93274 93375: gap of unknown length
* 93375 97602: contig of 4228 bp in length
* 97602 97603: gap of unknown length
* 97603 100909: contig of 3207 bp in length

* 100910 101009: gap of unknown length
* 101010 104900: contig of 3891 bp in length
* 104901 105000: gap of unknown length
* 105001 107283: contig of 2283 bp in length
* 107284 107383: gap of unknown length
* 107384 111626: contig of 4243 bp in length
* 111627 111726: gap of unknown length
* 111727 114910: contig of 3184 bp in length
* 114911 115010: gap of unknown length
* 115011 117629: contig of 2619 bp in length
* 117630 117729: gap of unknown length
* 117730 119406: contig of 1677 bp in length
* 119407 119506: gap of unknown length
* 119507 122616: contig of 3110 bp in length
* 122617 122716: gap of unknown length
* 122717 125803: contig of 3087 bp in length
* 125804 125903: gap of unknown length
* 125904 128384: contig of 2481 bp in length
* 128385 128484: gap of unknown length
* 128485 130846: contig of 2362 bp in length
* 130847 130946: gap of unknown length
* 130947 133969: contig of 3023 bp in length
* 133970 134069: gap of unknown length
* 134070 136593: contig of 2524 bp in length
* 136594 136693: gap of unknown length
* 136694 138664: contig of 2171 bp in length
* 138665 138964: gap of unknown length
* 138965 141786: contig of 2822 bp in length
* 141787 141886: gap of unknown length
* 141887 143704: contig of 1818 bp in length
* 143705 143804: gap of unknown length
* 143805 145781: contig of 1977 bp in length
* 145782 145881: gap of unknown length
* 145882 148147: contig of 2266 bp in length
* 148148 148247: gap of unknown length
* 148248 149829: contig of 1582 bp in length
* 149830 149929: gap of unknown length
* 149930 151551: contig of 1622 bp in length
* 151552 151651: gap of unknown length
* 151652 153613: contig of 1962 bp in length
* 153614 153713: gap of unknown length
* 153714 155443: contig of 1730 bp in length
* 155444 155543: gap of unknown length
* 155544 157807: contig of 2264 bp in length
* 157808 157907: gap of unknown length
* 157908 159583: contig of 1676 bp in length
* 159584 159683: gap of unknown length
* 159684 161415: contig of 1732 bp in length
* 161416 161515: gap of unknown length
* 161516 162970: contig of 1455 bp in length
* 162971 163070: gap of unknown length
* 163071 164093: contig of 1023 bp in length
* 164094 164193: gap of unknown length
* 164194 165392: contig of 1199 bp in length
* 165393 165492: gap of unknown length
* 165493 166326: contig of 1434 bp in length
* 166327 167026: gap of unknown length
* 167027 168102: contig of 1076 bp in length
* 168103 168202: gap of unknown length
* 168203 169537: contig of 1335 bp in length
* 169538 169637: gap of unknown length
* 169638 170962: contig of 1325 bp in length
* 170963 171062: gap of unknown length
* 171063 172084: contig of 1022 bp in length.

FEATURES

source

1. 172084
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-779P15"

BASE COUNT 51612 a 33289 c 31449 g 50869 t 4845 others
ORIGIN

Query Match 3.1%; Score 25; DB 76; Length 172084;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OR 770 aataataataataactaatg 794
Db 164559 AATPAATPAATPAATCAATG 164583

RESULT 7

AC073719/c AC073719 198057 bp DNA HTG 18-JUL-2000

LOCUS Mus musculus clone Rp23-208C14, WORKING DRAFT SEQUENCE, 12 ordered
DEFINITION pieces.

ACCESSION AC073719 GI:9256767

VERSION AC073719.2 GI:9256767

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 198057)

DOE Joint Genome Institute.

Sequencing of Mouse

Unpublished

2 (bases 1 to 198057)

DOE Joint Genome Institute.

Direct Submission

Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced GI:8810336.

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

Project Information

Center Project Name: 1816806

Center clone name: RPCI-23_208C14

Summary Statistics

Consensus quality: 190668 bases at least Q40

Consensus quality: 195558 bases at least Q30

Consensus quality: 196529 bases at least Q20

Estimated insert size: 213120; agarose-*fp* estimation

Quality coverage: 9.85 in Q20 bases; agarose-*fp* estimation

Quality coverage: 9.55 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 51188: contig of 51188 bp in length

51189 51288: gap of unknown length

51289 58807: contig of 7519 bp in length

58808 58907: gap of unknown length

58908 94961: contig of 36054 bp in length

94962 95061: gap of unknown length

95062 97717: contig of 2656 bp in length

97718 97817: gap of unknown length

97818 101072: contig of 3225 bp in length

101073 101172: gap of unknown length

101173 127195: contig of 26023 bp in length

127196 127295: gap of unknown length

127296 144047: contig of 16752 bp in length

144048 144147: gap of unknown length

144148 153335: contig of 9188 bp in length

153336 153435: gap of unknown length
 * 153436 183081: contig of 29646 bp in length
 * 183082 183181: gap of unknown length
 * 183182 184574: contig of 1393 bp in length
 * 184575 184674: gap of unknown length
 * 184675 187625: contig of 2951 bp in length
 * 187626 187725: gap of unknown length
 * 187726 198057: contig of 10332 bp in length.
 Location/Qualifiers
 1. 198057

FEATURES

source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-208C14"

BASE COUNT 44348 a 53180 c 52768 g 46659 t 1102 others
 ORIGIN

Query Match 3.1%; Score 25; DB 75; Length 198057;
 Best Local Similarity 100.0%; Pred. No. 0.0051;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 771 ataataataataactcaatgc 795
 ||||||||||||||||||||||||||||
 Db 125761 ATTAATAATAATAACTCAATGC 125737

RESULT 8
 AL357974 198777 bp DNA HTG 09-FEB-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-398M15, ** SEQUENCING IN
 DEFINITION PROGRESS ***, 2 unordered pieces.
 AL357974
 ACCESSION AL357974.1 GI:12657152
 VERSION HTG: HTGS.PHASE1; HTGS.ACTIVERIN; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 198777)
 Mashreghi-Mohammadi, M.
 Direct Submission
 Submitted (07-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Feb 2, 2001 this sequence version replaced gi:10241379.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: ba398M15
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 3% of reads
 Dye-terminator Big Dye; 96% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Consensus quality: 198106 bases at least Q40
 Consensus quality: 198261 bases at least Q30
 Consensus quality: 198387 bases at least Q20
 Insert size: 198677; sum-of-contigs
 Insert size: 195781; 3.8% error; agarose-fp
 Quality coverage: 13.00x in Q20 bases; sum-of-contigs Quality
 coverage: 13.77x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 168146: contig of 168146 bp in length
 * 168147 168246: gap of 100 bp
 * 168247 198777: contig of 30531 bp in length.
 Location/Qualifiers
 1. 198777

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-398M15"
 /clone_id="RPCT-11.2"
 1. 168146
 /note="assembly:fragment:06204
 fragment:chain:1"
 168247..198777
 /note="assembly:fragment:04558
 fragment:chain:1"

misc_feature

BASE COUNT 64964 a 34398 c 34202 g 65113 t 100 others
 ORIGIN

Query Match 3.1%; Score 25; DB 80; Length 198777;
 Best Local Similarity 100.0%; Pred. No. 0.0051;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 763 aacaagtaataataataataa 787
 ||||||||||||||||||||||||||||
 Db 161313 AACAGTAATAATAATAATAA 161289

RESULT 9
 AC073691 231443 bp DNA HTG 29-JUN-2000
 LOCUS Mus musculus clone RP23-144G23, WORKING_DRAFT SEQUENCE, 32
 DEFINITION unordered pieces.
 AC073691
 ACCESSION AC073691.1 GI:8810308
 VERSION HTG: HTGS.PHASE1; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 231443)
 DOE Joint Genome Institute.
 Sequencing of Mouse
 Unpublished
 2 (bases 1 to 231443)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 ----- Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov
 ----- Project Information
 Center Project Name: 1792335
 Center clone name: RPCT-23_144G23
 ----- Summary Statistics
 Consensus quality: 208474 bases at least Q40
 Consensus quality: 219397 bases at least Q30
 Consensus quality: 221920 bases at least Q20
 Estimated insert size: 179000; agarose-fp estimation
 Estimated insert size: 228343; sum-of-contigs estimation
 Quality coverage: 9.04 in Q20 bases; agarose-fp estimation
 Quality coverage: 7.09 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1045: contig of 1045 bp in length
1046 1145: gap of unknown length
1146 2377: contig of 1232 bp in length
2378 2477: gap of unknown length
2478 3674: contig of 1197 bp in length
3675 3774: gap of unknown length
3775 4946: contig of 1172 bp in length
4947 5046: gap of unknown length
5047 6054: contig of 1008 bp in length
6055 6155: gap of unknown length
6155 7485: contig of 1331 bp in length
7486 8676: contig of 1091 bp in length
8677 10069: contig of 1293 bp in length
10070 10169: gap of unknown length
10170 11272: contig of 1103 bp in length
11273 12909: contig of 1537 bp in length
12910 13009: gap of unknown length
13010 14675: contig of 1666 bp in length
14676 14775: gap of unknown length
14776 15879: contig of 1104 bp in length
15880 17559: gap of unknown length
17560 17659: gap of unknown length
17660 19068: contig of 1409 bp in length
19069 19168: gap of unknown length
19169 20444: contig of 1276 bp in length
20445 21648: gap of unknown length
21649 21748: contig of 1104 bp in length
21749 23422: contig of 1674 bp in length
23423 23522: gap of unknown length
23523 25715: contig of 2193 bp in length
25716 25815: gap of unknown length
25816 27924: contig of 2109 bp in length
27925 28024: gap of unknown length
28025 31025: contig of 3001 bp in length
31026 31125: gap of unknown length
31126 34688: contig of 3563 bp in length
34689 34788: gap of unknown length
34789 41906: contig of 7118 bp in length
41907 42006: gap of unknown length
42007 47079: contig of 5073 bp in length
47080 47179: gap of unknown length
47180 56114: contig of 8935 bp in length
56115 56214: gap of unknown length
56215 66066: contig of 9852 bp in length
66067 66166: gap of unknown length
66167 78095: contig of 11929 bp in length
78096 78195: gap of unknown length
78196 93045: contig of 14850 bp in length
93046 93145: gap of unknown length
93146 108338: contig of 15193 bp in length
108339 108438: gap of unknown length
108439 128434: contig of 19996 bp in length
128435 128534: gap of unknown length
128535 145880: contig of 17346 bp in length
145881 145980: gap of unknown length
145981 173500: contig of 27520 bp in length
173501 173600: gap of unknown length
173601 231443: contig of 57843 bp in length.

```

FEATURES
 Source
 1. 231443
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-144G23"

BASE COUNT 52186 a 61071 c 61200 g 53732 t 3254 others
 ORIGIN

Query Match 3.1%; Score 25; DB 75; Length 231443;
 Best Local Similarity 100.0%; Pred. No. 0.0051;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Y 771 ataataataataactcaatgc 795
 |||
 Db 29245 ATAAATAATAATAACTCAATGC 29269

RESULT 10

AC073811 246962 bp DNA HTG 18-JUL-2000
 LOCUS
 DEFINITION Mus musculus clone RP23-61D17, WORKING DRAFT SEQUENCE, 23 ordered
 pieces.
 AC073811.2 GI:9256810
 AC073811
 VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 2 (bases 1 to 246962)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Jul 18, 2000 this sequence version replaced gi:8810428.

-----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

Project Information
 Center Project Name: 1760385
 Center Clone name: RPC1-23_61D17

Summary Statistics
 Consensus quality: 235325 bases at least Q40
 Consensus quality: 242760 bases at least Q30
 Consensus quality: 244123 bases at least Q20
 Estimated insert size: 247000; agarose-1p estimation
 Estimated insert size: 245912; sum-of-contigs estimation
 Quality coverage: 7.33 in Q20 bases; agarose-1p estimation
 Quality coverage: 7.36 in Q20 bases; sum-of-contigs estimation.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

```

1 2990 3089: contig of 2989 bp in length
3090 6703: contig of 3614 bp in length
6704 6803: gap of unknown length
6804 26752: contig of 19949 bp in length
26753 26853: gap of unknown length
26853 34845: contig of 7993 bp in length
34846 34945: gap of unknown length
34946 49914: contig of 14969 bp in length
49915 50014: gap of unknown length
50015 61588: contig of 11574 bp in length

```



```
repeat_region      8936. .9057
/note="MIR repeat: matches 71. .197 of consensus"
repeat_region      9069. .9153
/note="5 copies 17 mer 71% conserved"
repeat_region      9589. .9780
/note="96 copies 2 mer tc 70% conserved"
repeat_region      9609. .9780
/note="43 copies 4 mer ttc 73% conserved"
repeat_region      9614. .9773
/note="10 copies 16 mer 75% conserved"
repeat_region      9619. .9762
/note="12 copies 12 mer 77% conserved"
misc_feature       complement(961. .10266)
/note="match: GSS: Em:AQ366486"
misc_feature       complement(9807. .10272)
/note="match: GSS: Em:AQ174667"
misc_feature       10295. .10789
/note="match: GSS: Em:AQ473771"
repeat_region      10485. .10791
/note="AluX repeat: matches 1. .304 of consensus"
misc_feature       complement(10843. .11112)
/note="match: GSS: Em:B44529"
repeat_region      11307. .11423
/note="L2 repeat: matches 2588. .2710 of consensus"
repeat_region      11526. .11717
/note="MIR repeat: matches 31. .233 of consensus"
repeat_region      11768. .11781
/note="L1P83 repeat: matches 5654. .5666 of consensus"
repeat_region      11909. .11919
/note="L1P83 repeat: matches 5666. .5789 of consensus"
repeat_region      11984. .12348
/note="L1P83 repeat: matches 5782. .6147 of consensus"
misc_feature       13204. .13606
/note="match: GSS: Em:AQ09872"
misc_feature       13204. .13605
/note="match: GSS: Em:AQ010332"
repeat_region      13456. .13467
/note="MIR repeat: matches 57. .68 of consensus"
repeat_region      13468. .13772
/note="AluSP repeat: matches 1. .310 of consensus"
repeat_region      13773. .13855
/note="MIR repeat: matches 68. .149 of consensus"
repeat_region      14174. .14462
/note="AluX repeat: matches 1. .290 of consensus"
repeat_region      14565. .14873
/note="AluB repeat: matches 1. .308 of consensus"
repeat_region      15277. .15343
/note="MIR repeat: matches 81. .147 of consensus"
repeat_region      15665. .15763
/note="3 copies 33 mer 82% conserved"
repeat_region      15783. .15870
/note="2 copies 44 mer 90% conserved"
repeat_region      16734. .16800
/note="AluY/FLAM repeat: matches 23. .89 of consensus"
repeat_region      16852. .17148
/note="AluSg repeat: matches 1. .295 of consensus"
repeat_region      17200. .17649
/note="match: GSS: Em:AQ163770"
misc_feature       17259. .17394
/note="MIR repeat: matches 107. .241 of consensus"
repeat_region      18947. .19175
/note="L2 repeat: matches 2495. .2750 of consensus"
repeat_region      19176. .19352
/note="MER5B repeat: matches 1. .178 of consensus"
repeat_region      19375. .19488
/note="L2 repeat: matches 2304. .2421 of consensus"
repeat_region      19700. .20015
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region      20204. .20300
/note="L2 repeat: matches 2593. .2708 of consensus"
repeat_region      21153. .21217
/note="MIR repeat: matches 112. .176 of consensus"
repeat_region      21648. .21681
```

```
repeat_region      /note="17 copies 2 mer ac 82% conserved"
21824. .22005
/note="MER91A repeat: matches 6. .196 of consensus"
repeat_region      22118. .22422
/note="AluX repeat: matches 1. .305 of consensus"
misc_feature       complement(23992. .24251)
/note="match: GSS: Em:AQ633877"
misc_feature       24276. .24717
/note="match: GSS: Em:AQ450746"
misc_feature       24276. .24713
/note="match: GSS: Em:AQ334737"
repeat_region      25148. .25243
/note="24 copies 4 mer 99% 75% conserved"
repeat_region      25149. .25216
/note="34 copies 2 mer ag 75% conserved"
repeat_region      25157. .25240
/note="7 copies 12 mer 78% conserved"
repeat_region      25162. .25241
/note="5 copies 16 mer 82% conserved"
repeat_region      25194. .25292
/note="3 copies 33 mer 79% conserved"
repeat_region      25196. .25280
/note="5 copies 17 mer 72% conserved"
repeat_region      25685. .25825
/note="MIR repeat: matches 8. .179 of consensus"
repeat_region      25826. .26114
/note="AluSg repeat: matches 1. .288 of consensus"
repeat_region      26115. .26189
/note="MIR repeat: matches 179. .245 of consensus"
repeat_region      26351. .26472
/note="L2 repeat: matches 2377. .2710 of consensus"
repeat_region      26771. .27231
/note="MIR1G repeat: matches 74. .512 of consensus"
repeat_region      27332. .27574
/note="L2 repeat: matches 2042. .2297 of consensus"
misc_feature       complement(27558. .27964)
```

```
Query Match      3.0%; Score 24; DB 89; Length 144818;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      770 aataataataataactaat 793
DB      4998 AATAATAATAATAACTAAT 4975
```

```
RESULT  12
AC013396      155986 bp      DNA      16-MAR-2000
LOCUS      Homo sapiens chromosome 2 clone RP11-489G24 map 2, WORKING DRAFT
DEFINITION      AC013396
ACCESSION      AC013396.4 GI:7107974
VERSION      HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 155986)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 2, clone RP11-489G24
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 155986)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguski,K., Boukhvalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deailellano,K., Dewar,K., Domino,M., Donahue,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGuff,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
```


1	1085:	contig of 1085	bp in length
2	1086:	contig of 1086	bp in length
3	1185:	gap of unknown length	
4	1186:	contig of 1186	bp in length
5	1255:	contig of 1255	bp in length
6	2454:	gap of unknown length	
7	3597:	contig of 1244	bp in length
8	3798:	gap of unknown length	
9	3798:	contig of 1006	bp in length
10	4403:	gap of unknown length	
11	4904:	contig of 1970	bp in length
12	6873:	gap of unknown length	
13	6874:	contig of 2538	bp in length
14	6974:	contig of 2538	bp in length
15	9512:	gap of unknown length	
16	9512:	contig of 1288	bp in length
17	9612:	contig of 1288	bp in length
18	10910:	gap of unknown length	
19	12902:	contig of 1893	bp in length
20	12903:	gap of unknown length	
21	13002:	gap of unknown length	
22	13003:	contig of 1397	bp in length
23	14000:	gap of unknown length	
24	14499:	gap of unknown length	
25	14500:	contig of 1959	bp in length
26	14500:	gap of unknown length	
27	16559:	contig of 1750	bp in length
28	16559:	contig of 1750	bp in length
29	18309:	gap of unknown length	
30	18309:	contig of 2544	bp in length
31	20953:	contig of 2544	bp in length
32	20953:	gap of unknown length	
33	23674:	contig of 2632	bp in length
34	23674:	contig of 2632	bp in length
35	23775:	gap of unknown length	
36	23775:	contig of 2397	bp in length
37	262172:	gap of unknown length	
38	262172:	contig of 2910	bp in length
39	262172:	contig of 2910	bp in length
40	29182:	gap of unknown length	
41	29182:	contig of 2661	bp in length
42	31963:	contig of 2661	bp in length
43	31963:	gap of unknown length	
44	32063:	contig of 2662	bp in length
45	34745:	contig of 2662	bp in length
46	34745:	gap of unknown length	
47	34845:	contig of 2917	bp in length
48	34845:	contig of 2917	bp in length
49	37762:	gap of unknown length	
50	37762:	contig of 2947	bp in length
51	40809:	contig of 2947	bp in length
52	40809:	gap of unknown length	
53	44229:	contig of 3821	bp in length
54	44229:	contig of 3821	bp in length
55	44730:	gap of unknown length	
56	44730:	contig of 2792	bp in length
57	44830:	contig of 2792	bp in length
58	47622:	gap of unknown length	
59	47622:	contig of 2795	bp in length
60	47722:	contig of 2795	bp in length
61	50517:	gap of unknown length	
62	50517:	contig of 4126	bp in length
63	54743:	gap of unknown length	
64	54743:	contig of 2978	bp in length
65	54843:	contig of 2978	bp in length
66	57921:	gap of unknown length	
67	57921:	contig of 3366	bp in length
68	61386:	gap of unknown length	
69	61386:	contig of 3856	bp in length
70	61387:	contig of 3856	bp in length
71	65243:	gap of unknown length	
72	65243:	contig of 3273	bp in length
73	66116:	gap of unknown length	
74	66116:	contig of 3183	bp in length
75	68716:	gap of unknown length	
76	71899:	contig of 3952	bp in length
77	71899:	contig of 3952	bp in length
78	75951:	gap of unknown length	
79	75951:	contig of 4749	bp in length
80	80809:	gap of unknown length	
81	80809:	contig of 4898	bp in length
82	85798:	gap of unknown length	
83	85798:	contig of 5003	bp in length
84	90901:	gap of unknown length	
85	90901:	contig of 6075	bp in length
86	97076:	gap of unknown length	
87	97076:	contig of 6507	bp in length
88	97176:	gap of unknown length	
89	97176:	contig of 6507	bp in length
90	103682:	gap of unknown length	
91	103682:	contig of 6507	bp in length
92	103782:	gap of unknown length	

	*	103783	110814:	contig of 7032 bp in length
	*	110815	110914:	gap of unknown length
	*	110915	115167:	contig of 4233 bp in length
	*	115168	115267:	gap of unknown length
	*	115268	124863:	contig of 9596 bp in length
	*	124864	124963:	gap of unknown length
	*	124864	133477:	contig of 8474 bp in length
	*	133438	133577:	gap of unknown length
	*	133538	147665:	contig of 14128 bp in length
	*	147666	147765:	gap of unknown length
	*	147766	162486:	contig of 14721 bp in length
URS				
source			Location/Qualifiers	
			1..162486	
			/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/chromosome="10"	
			/clone="RP1-105119"	
			/clone_1fb="RPc1-11"	
			1..1085	
misc_feature			/note="assembly_name:Contig42"	
misc_feature			1186..2253	
			/note="assembly_name:Contig40"	
misc_feature			2454..3697	
			/note="assembly_name:Contig20"	
misc_feature			3798..4803	
			/note="assembly_name:Contig26"	
misc_feature			4904..6873	
			/note="assembly_name:Contig28"	
misc_feature			6974..9511	
			/note="assembly_name:Contig29"	
misc_feature			9612..10909	
			/note="assembly_name:Contig30"	
misc_feature			10101..12902	
			/note="assembly_name:Contig31"	
misc_feature			13003..14399	
			/note="assembly_name:Contig33"	
misc_feature			14500..16458	
			/note="assembly_name:Contig34"	
misc_feature			16559..18308	
			/note="assembly_name:Contig35"	
misc_feature			18409..20952	
			/note="assembly_name:Contig36"	
misc_feature			21053..23674	
			/note="assembly_name:Contig37"	
misc_feature			23775..26171	
			/note="assembly_name:Contig38"	
misc_feature			26272..29181	
			/note="assembly_name:Contig39"	
misc_feature			29282..31962	
			/note="assembly_name:Contig40"	
misc_feature			32063..34744	
			/note="assembly_name:Contig42"	
misc_feature			34845..37761	
			/note="assembly_name:Contig43"	
misc_feature			37862..40808	
			/note="assembly_name:Contig44"	
misc_feature			40909..44729	
			/note="assembly_name:Contig45"	
misc_feature			44830..47621	
			/note="assembly_name:Contig46"	
misc_feature			47722..50516	
			/note="assembly_name:Contig47"	
misc_feature			50617..54742	
			/note="assembly_name:Contig49"	
misc_feature			54843..57820	
			/note="assembly_name:Contig50"	
misc_feature			57921..61286	
			/note="assembly_name:Contig51"	
misc_feature			61387..65242	
			/note="assembly_name:Contig52"	
misc_feature			65343..68615	
			/note="assembly_name:Contig53"	
misc_feature			68716..71898	

misc_feature /note="assembly_name:Contig54"
71999..75950
/note="assembly_name:Contig55"
misc_feature /note="assembly_name:Contig55"
76051..80799
/note="assembly_name:Contig56"
misc_feature /note="assembly_name:Contig56"
80900..85797
/note="assembly_name:Contig57"
misc_feature /note="assembly_name:Contig57"
85898..90900
/note="assembly_name:Contig58"
misc_feature /note="assembly_name:Contig58"
91001..97075
/note="assembly_name:Contig59"
misc_feature /note="assembly_name:Contig59"
97176..103682
/note="assembly_name:Contig60"
Query Match 3.0%; Score 24; DB 72; Length 162486;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 770 aataaataaataaactcaat 793
|||||
Db 138980 AATAAATAAATAAATACTCAAT 139003
RESULT 14
MMU278435/c
LOCUS MMU278435 192519 bp DNA ROD 01-DEC-2000
DEFINITION Mus musculus Ranbp7 gene, Staf gene and Weel gene.
ACCESSION A0278435
VERSION A0278435.1 GI:11342590
KEYWORDS Importin; RANBP7 gene; selenocysteine tRNA gene transcription
activating factor; staf gene; Weel gene; Weel protein.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192519)
Cichutek, A., Winterpacht, A., Hankeln, T., Schmidt, E.R. and
Zabel, B.O.
Mouse PAC clone RPCI711N1025603, sequenced INDRGP project
(Comparative Sequencing of a 1 Mb Region in Man (Chromosome 11p15)
and Mouse (Chromosome 7))
Unpublished
2 (bases 1 to 192519)
Cichutek, A.
Direct Submission
Submitted (13-JUN-2000) Cichutek A., Kinderklinik Uniklinik Mainz,
Molekulargenetisches Labor, Obere Zahlbacher Str. 63, 55131 Mainz,
GERMANY
revised by {3}
3 (bases 1 to 192519)
Cichutek, A.
Direct Submission
Submitted (22-NOV-2000) Cichutek A., Kinderklinik Uniklinik Mainz,
Molekulargenetisches Labor, Obere Zahlbacher Str. 63, 55131 Mainz,
GERMANY
FEATURES
Source Location/Qualifiers
1..192519
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7"
/clone="PAC clone RPCI711N1025603"
/note="complete PAC clone sequence"
join(4153..4341,6643..26727,27811..27945,28616..28715,
30239..30318,30448..30564,31992..32081,32508..32673,
32817..32977,33959..34087,34624..34690,34775..34900,
35426..35523,36610..36705,36843..37063,37152..37357,
38541..38747,39830..39946,40847..40944)
/gene="Ranbp7"
4153..4341
/gene="Ranbp7"
/number=1
join(4258..4341,12902..12983,15522..15675,17438..17596,
22774..22930,24994..25083,26436..26530,26643..26727,
27774..22930,24994..25083,26436..26530,26643..26727,

27811..27945,28616..28715,30239..30318,30448..30564,
31992..32081,32508..32673,32817..32977,33959..34087,
34624..34690,34775..34900,35426..35523,36610..36705,
36843..37063,37152..37357,38541..38747,39830..39946,
40847..40944)
/gene="Ranbp7"
/function="Importin 7/Importin beta heterodimer mediates
import of histone H1"
/codon_start=1
/product="Ranbp7/Importin 7"
/protein_id="CAC17143.1"
/db_xref="GI:11342591"
/translation="MDPNTIEALRGMDPALREAROLNEAKHSNPNSTLIQTM
SEGLDLPVRQAGVITYLKNMITYPMDREAPGDIAPPTTEEDRHCIKREIVAHIS
PELIRVOLTTCIHIIKHDPSPMTALVNDIGFYLDSDNSACWGLIACGYQVKNK
YKRPDESPPLVAAHQHFLPYLKDREFIOLDSQSVLIQROIKFIKLYALVQYLP
LINOQNLTEWVEIKTVNNDVPRETIQVEDDRPELPMKCKKMAHIIARFERG
SPGNVSEKVEFEAVFLKFAVGOVLKLVLYQKROYMAPRVLOQTNTYINQVS
HATYKRNKPHIGIITDVFPLMCTYDADBEIQEDPXYIRKREDFEDFISPTTA
AQTLETFACSKREEVVQKTWGFQYQITLPEPADPKKDALHMTGLABILKRLKIY
KDQMEYMLQNHVEPLFSSSELYMRARACWVLAHCEVKEFSQNLQALBELTRCLID
DREMPYKVEAIALQVLAISNOEKAEXITPFIIPVQALHIIIRETENDDLTVNIQK
ICEYSEEPPIAIVEMTOHLAMTFENOVYIQGDEGSDKRAVTAMGLINTIDTLLSYE
DKKEITQQLGICLOVYIGTVLQCHVLEFEIETSLASHSLTCCQVSPOMQLDPLVEY
FOODGFDTDMPLLHNYVYVDITLSTQTKLEMTYSCKKVLGVGAGEDACAKA
KILEVILIQCKGIDQCIPLFEVALERLTREVKSELRTMQLQVLAIALVNPML
LNTLEMLREPNNEPVYTNHITQMLNDVDFGLHDKMKVGLICALIDMEQIPOVNL
QVSGQILPFIILFNGLRAYACHAEHNSDDDEDDDETEELSGSDIDDEDO
EYELIILAKMGDEGDEDEDEDAEETALEGSTIITIDDENPVDEYOIFAITQTON
RNWVYQALSHGMEORQKLODIATLADRRAAHESKMLEKGGIKFSPVPPSPFN
EGGPAPQMN
6643..26727
/gene="Ranbp7"
12902..12983
/number=8
/gene="Ranbp7"
15522..15675
/number=2
/gene="Ranbp7"
17438..17596
/number=3
/gene="Ranbp7"
/number=4
join(21456..121541,124967..125137,125380..125525,
127774..127951)
/gene="Weel"
21456..121541
/gene="Weel"
/number=8
22774..22930
/number=5
/gene="Ranbp7"
24994..25083
/number=5
/gene="Ranbp7"
26436..26530
/number=6
/gene="Ranbp7"
27811..27945
/number=7
/gene="Ranbp7"
/number=9
28616..28715
/number=10
/gene="Ranbp7"
30239..30318
/number=10
/gene="Ranbp7"
/number=11
30448..30564
/number=12
/gene="Ranbp7"
31992..32081
/number=13

```

exon      32508..32673
           /gene="Ranbp7"
           /number=14
exon      32817..32977
           /gene="Ranbp7"
           /number=15
exon      33959..34087
           /gene="Ranbp7"
           /number=16
exon      34624..34690
           /gene="Ranbp7"
           /number=17
exon      34775..34900
           /gene="Ranbp7"
           /number=18
exon      35426..35523
           /gene="Ranbp7"
           /number=19
exon      36610..36705
           /gene="Ranbp7"
           /number=20
exon      36843..37063
           /gene="Ranbp7"
           /number=21
exon      37152..37357
           /gene="Ranbp7"
           /number=22
exon      38541..38747
           /gene="Ranbp7"
           /number=23
exon      39630..39946
           /gene="Ranbp7"
           /number=24
exon      40847..40944
           /gene="Ranbp7"
           /number=25
exon      5295..35424
           /gene="Staf"
           /number=1
gene       join(55295..55424,56300..56392,57194..57277,57941..58024,
           59890..60086,61463..61537,62879..62998,66090..66165,
           67614..67739,69013..69192,71935..72162,74550..74698,
           77587..77748,78892..79038,80113..80358)
           /gene="Staf"
           join(55349..55424,56300..56392,57194..57277,57941..58024,
           59890..60086,61463..61537,62879..62998,66090..66165,
           67614..67739,69013..69192,71935..72162,74550..74698,
           77587..77748,78892..79038,80113..80358)
           /gene="Staf"
           /codon_start=1
           /product="selenocysteine tRNA gene transcription
activating factor"
           /protein_id="CAC17144.1"
           /db_xref="GI:11342592"
           /translation="MTEFPGGMEADQHVLTCLTEAVTVADGDLNENGVSLQAVTLA
DGSFAYIHNHSGKRLDQVIOLEDGSAAYVHVPIPKSTGSLRLEDQAVOLEDG
TFAFIHNTSKDQSLQAVLEDGTTAYIHNAVYQSDTILAIADQVAVLHG
DATIDPTISALFOYAKVSDIGSDGVTSTGEMGENOEKMOIVLOGHTRVTKSG
OSGEKAFRCYVDCGKLYTFAHILKVAHERSHTDPRPCQSGCGKAPAGYGLKSF
RTTGKEPRYRCSEDNCTKSKFTSGDLQKHRTTGERPKPIEGCCRSRTTSIRKY
HIRTGGERPYCTEPGCGAFASATYKNHVIHTGERPYCTVPCDRTFTEYSSL
YKHNVHTSKPYNCNMGCTYKOISLANKKRAHNDTEPIEEDQAFEPFGQGD
DYLKGSQITVTVGDEDIYVQAVTVOGLSQOYTLISDQGHQNIQAADQAG
NTLTMTQDGTPLTVPFHADVISAGTSHVAMVTAEGTEGOVAIVODLAAFTTAS
EMGHQSHHLVTTETRRPLTVAFTSNCTQIADVQGEPRPSLEAIRIRASIQGERTPL
DP"
exon      56300..56392
           /gene="Staf"
           /number=2
exon      57194..57277
           /gene="Staf"
           /number=3
exon      57941..58024

```

```

exon      /gene="Staf"
           /number=4
           59890..60086
           /gene="Staf"
           /number=5
           61463..61537

Query Match      3.0%; Score 24; DB 94; Length 192519;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      764  acaagtaataataataataa 787
           |||||
Db 102198  ACAAGTATATAATAATAATAA 102175

```

```

RESULT 15
AC024694
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC024694 215516 bp DNA HTG 01-MAR-2000
Mus musculus clone RP23-189B10, *** SEQUENCING IN PROGRESS ***, 20
unordered pieces.
AC024694
AC024694.1 GI:7139921
HTG; HTGS_PHASE1.
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.
1 (bases 1 to 215516)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-189B10
Unpublished
2 (bases 1 to 215516)
Anderson,S., Baldwin,J., Barra,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukngalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Gallagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Hottel,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Lavigne,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivar,T.M., Oliver,T., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L884
Center clone name: 189_B_10

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 2903: contig of 2903 bp in length

1
2904 3003: gap of 100 bp
3004 6543: contig of 3540 bp in length
6544 6643: gap of 100 bp
6644 11328: contig of 4685 bp in length
11329 11428: gap of 100 bp
11429 14900: contig of 3472 bp in length
14901 15000: gap of 100 bp
15001 19603: contig of 4603 bp in length
19604 19703: gap of 100 bp
19704 25033: contig of 5330 bp in length
25034 25133: gap of 100 bp
25134 31568: contig of 6435 bp in length
31569 31668: gap of 100 bp
31669 40823: contig of 9155 bp in length
40824 40923: gap of 100 bp
40924 48607: contig of 7684 bp in length
48608 48707: gap of 100 bp
48708 55161: contig of 6454 bp in length
55162 55261: gap of 100 bp
55262 63685: contig of 8424 bp in length
63686 63785: gap of 100 bp
63786 72355: contig of 8570 bp in length
72356 72455: gap of 100 bp
72456 80220: contig of 7765 bp in length
80221 80320: gap of 100 bp
80321 91780: contig of 11460 bp in length
91781 91880: gap of 100 bp
91881 105628: contig of 13748 bp in length
105629 105728: gap of 100 bp
105729 120816: contig of 15088 bp in length
120817 120916: gap of 100 bp
120917 136141: contig of 15225 bp in length
136142 136241: gap of 100 bp
136242 150918: contig of 14677 bp in length
150919 151018: gap of 100 bp
151019 175203: contig of 24185 bp in length
175204 175303: gap of 100 bp
175304 215516: contig of 40213 bp in length.

FEATURES

source

1. 215516
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone.lib="RP23-189B10"

BASE COUNT 55241 a 54375 c 53217 g 49260 t 3423 others

ORIGIN

Query Match

Best Local Similarity 3.0%; Score 24; DB 69; Length 215516;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 764 acaagtaataataataataa 787
|||||

DB 49013 ACNAGTAATAATAATAATAA 49036

Search completed: June 3, 2001, 04:42:58
Job time: 5104 sec


```

51 AsnleuysgLyHsHsValaArgLeuCyLysProCysLysLeuGluPr 67
|||||
205 AACCTGAAGGACATCATGTGAGCTCTGTAAACCATGCAAGCTTGAGCC 254
67 oGluPrArgLeuTrpValaValProGlyAlaLeuProGlnVal 81
|||||
255 AGAGCCCCCGCTTTGGGTGGTGGCTGGGCGACATCCACAGAGTG 297

seq_name: gb_est7:AA422178

seq_documentation_block:
LOCUS AA422178 503 bp mRNA EST 16-OCT-1997
DEFINITION zv31907.t1 Soares ovary tumor NBHOT Homo sapiens cDNA clone
IMAGE:755292 5', mRNA sequence.
ACCESSION AA422178
VERSION AA422178.1 GI:2101029
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 503)
AUTHORS Hillier,L., Allen,M., Bowles,J., Dubuque,T., Giesel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Martin,M., Martin,D., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie
,T., Waterston,R. and Wilson,R.
MashU-Merck EST Project 1997
UNPUBLISHED (1997)
JOURNAL
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: eastewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28n13 rev2 ER from Amersham
High quality sequence stop: 503.
FEATURES
Source
Location/Qualifiers
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:755292"
/clone_lib="Soares ovary tumor NBHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pTR73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTCACATCTGAGTGGGAGCGCGCGCGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pTR73 vector
(Pharmacia). Library constructed by Berto Soares and
M.Fatima Bonaldo."
BASE COUNT 108 a 167 c 119 g 109 t
ORIGIN
alignment_scores:
Quality: 418.00 Length: 76
Ratio: 5.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-599-087-5 x AA422178
Align seg 1/1 to: AA422178 from: 1 to: 503
|||||
1 MetArgLeuValLeuSerSerLeuLeuCyLysLeuLeuLeuCySpH 17
|||||

```

```

16 ATGAGCCTTCTAGTCCTTTCCAGCCTGCTGTATACCTGCTTCTGCT 65
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerG 34
|||||
66 CTCATCTTCTTCACAGAGGAGGAGAGGCGCTCTCCAGAGCTGTGACG 115
34 lYArgArgThrArgLeuCyCysHsArgValProSerProAsnSerThr 50
|||||
116 GCAGGAGAACCGAGGCTGTGCTCCACCGAGTGCTTACGCCCAACTCAAC 165
51 AsnleuysgLyHsHsValaArgLeuCyLysProCysLysLeuGluPr 67
|||||
166 AACCTGAAGGACATCATGTGAGCTCTGTAAACCATGCAAGCTTGAGCC 215
67 oGluPrArgLeuTrpValaValProGly 76
|||||
216 AGAGCCCCCGCTTTGGGTGGTGGCTGGGCGACATCCACAGAGTG 243

seq_name: gb_est80:BE899580

seq_documentation_block:
LOCUS BE899580 906 bp mRNA EST 29-SEP-2000
DEFINITION 601682443F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3952523 5',
mRNA sequence.
ACCESSION BE899580
VERSION BE899580.1 GI:10367234
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 906)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
NATIONAL INSTITUTE OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL
UNPUBLISHED (1999)
CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: LICM822 row: f column: 12
High quality sequence start: 23
High quality sequence stop: 775.
FEATURES
Source
Location/Qualifiers
1..906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3952523"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGCG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 219 a 282 c 219 g 186 t
ORIGIN
alignment_scores:
Quality: 416.00 Length: 83
Ratio: 5.136 Gaps: 2
Percent Similarity: 97.590 Percent Identity: 97.590

alignment_block:

```


1 (Pages 1 to 342)
Morra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisels, S., Kuchala, T., Lacy, M., Le, M., Martin, J., Morris, M., Neilsen, B., K., Steptoe, M., Tan, F., Underwood, C., Moore, B., Moelling, B., Wyle, T., Lennon, G., Soares, B., Wilson, R.

LOCUS	AA840147	436 bp	mRNA
-------	----------	--------	------

LOCUS	436 bp	mRNA	27-FEB-1998
DEFINITION	u01f07.r1 Soares_NMPu Mus musculus cDNA clone IMAGE:1433893 5'	EST	